

# Influencia y recomendación en redes sociales

---

Memoria del Proyecto Fin de Máster

Ana Cristina Mateos de Miguel



**Tutores:** Pablo Castells Azpilicueta

Simone Santini

**Curso:** 2012/13

**Fecha:** Septiembre 2013

## **Agradecimientos**

Quiero agradecer a mis tutores Pablo Castells y Simone Santini el esfuerzo y el tiempo invertido.

A David, por el cariño y el apoyo que siempre he recibido de ti y que me ha dado las fuerzas necesarias para continuar.

A mi familia, por su infinita paciencia, su cariño y comprensión.

## Contenido

Agradecimientos .....	2
Contenido .....	3
Tabla de ilustraciones.....	5
1. Introducción .....	6
1.1 Motivación .....	6
1.2 Objetivos .....	7
2. Estudio de las redes.....	9
2.1 Contexto histórico .....	9
2.2 Estado actual .....	11
2.3 Propiedades estructurales de las redes .....	12
Métricas locales.....	13
Métricas globales .....	14
2.4 Modelos de formación de redes .....	16
Modelo Erdős-Rényi .....	16
Enlace preferente: Barabási - Albert .....	17
3. Procesos en las redes .....	20
3.1 Percolación .....	20
3.2 Sistemas dinámicos en redes .....	22
Epidemias en redes .....	22
Propagación en redes.....	23
4. Propagación e influencia en las redes .....	25
4.1 Aspectos determinantes .....	25
Propagación y topología.....	25
Propagación y los enlaces débiles .....	25
4.2 Influencia en las redes sociales .....	26
Homofilia .....	27
Condiciones externas .....	27
Influencia social.....	27
4.3 Modelando la difusión en las redes .....	28
Sociología .....	28
Modelos físicos.....	29
4.4 Influencia social y sus aplicaciones .....	31

Marketing viral .....	31
Recomendación personalizada .....	33
5. Descripción del modelo.....	35
5.1 Estructura social .....	35
5.2 El modelo de opinión .....	35
5.3 El modelo de interacción.....	36
5.4 Dinámica de propagación.....	37
5.5 Metodología de la recomendación .....	39
5.6 Trabajo relacionado .....	40
6. Estudio experimental .....	42
6.1 Propósito .....	42
6.2 Descripción de las pruebas.....	42
6.3 Variables estudiadas .....	42
Usuarios por ítem preferido.....	43
Porcentaje de convergencia.....	43
Porcentaje de usuarios en ítem más popular .....	44
Tiempo de convergencia .....	44
Número medio de ítems preferidos.....	44
6.4 Arquitectura del programa.....	44
6.5 Análisis cualitativo.....	46
Preferencia por el mismo ítem.....	46
Preferencia por dos ítems .....	47
No convergencia.....	49
Tendencias por comunidades .....	50
6.6 Casos de estudio.....	51
Caso 1: Efecto de la densidad.....	51
Caso 2: Efecto de la influencia exógena .....	54
Caso 3: Efecto de la convicción .....	57
Caso 4: Introducción de un agente recomendador.....	60
7. Conclusiones.....	64
8. Trabajo futuro .....	65
9. Bibliografía .....	67

## Tabla de ilustraciones

Figura 1 - Hitos en el desarrollo de las redes sociales online entre 1997 y 2006 (Boyd & Ellison, 2008) .....	11
Figura 2 - Grafo dirigido vs no dirigido .....	12
Figura 3 - Ejemplo del grado de un nodo .....	13
Figura 4 - Cierre triádico .....	13
Figura 5 - Componente gigante en una red.....	16
Figura 6 - Distribución binomial del grado de los nodos (ER).....	17
Figura 7 - Modelo aleatorio vs enlace preferente .....	18
Figura 8 - Distribución power law del grado en un grafo Barabási-Albert de 4000 nodos con grado promedio 40.....	19
Figura 9 - Ejemplo de percolación (Newman M. , 2010) .....	21
Figura 10 - Transiciones permitidas en el modelo SI.....	22
Figura 11 - Transiciones permitidas en el modelo SIR.....	22
Figura 12 - Transiciones permitidas en el modelo SIS .....	23
Figura 13 - Transiciones permitidas en el modelo SIRS.....	23
Figura 14 - Ejemplo de red que favorece la propagación de un rumor.....	25
Figura 15 - Enlaces fuertes y débiles. (Kleinberg & Easley, 2010) .....	26
Figura 16 - Variación de la opinión media frente a $\lambda$ (Modelo LCCC) .....	30
Figura 17 - Normalización de las opiniones.....	36
Figura 18 - Pseudocódigo de la dinámica de propagación .....	38
Figura 19 - Secuencia del algoritmo .....	39
Figura 20 - Esquema de la aplicación .....	45
Figura 21 - Ejemplo de grafo visualizado con Mason .....	45
Figura 22 - Opiniones de un usuario.....	46
Figura 23 - Evolución preferencia por un ítem .....	46
Figura 24 - Preferencia de ítem por usuario (caso convergencia 1 ítem) .....	47
Figura 25 - Evolución preferencia 2 ítems .....	48
Figura 26- Preferencia de ítem por usuario (caso convergencia a 2 ítems) .....	48
Figura 27 - Evolución no convergencia.....	49
Figura 28 - Preferencia de ítem por usuario (caso no convergencia).....	49
Figura 29 - Evolución de la opinión en un grafo real .....	50
Figura 30 - Efectos observados al variar el grado promedio (Erdős-Rényi) .....	51
Figura 31 - Número de ítems preferidos al aumentar la densidad (Erdős-Rényi).....	52
Figura 32 - Efectos observados al variar el grado (Barabási-Albert) .....	53
Figura 33 - Número de ítems preferidos al variar la densidad (Barabási-Albert) .....	53
Figura 34 - Efectos observados al variar el parámetro $\alpha$ (Erdős-Rényi) .....	54
Figura 35 - Efectos observados al variar el parámetro $\alpha$ (Barabási-Albert).....	55
Figura 36 - Efectos observados al variar el parámetro $\alpha$ (grafo real) .....	56
Figura 37 -Número de ítems preferidos al variar la influencia exógena .....	57
Figura 38 - Efectos observados al variar la convicción (Erdős-Rényi) .....	58
Figura 39 - Efectos observados al variar la convicción (Barabási-Albert).....	59
Figura 40 - Efectos observados al variar la convicción (grafo real) .....	59
Figura 41 - Número de ítems preferidos al variar la convicción.....	60
Figura 42 - Evolución de la formación de opinión con recomendación .....	61
Figura 43 - Efecto del recomendador en función de la influencia exógena.....	62
Figura 44 - Efecto del recomendador con valor alto de influencia exógena.....	63

## 1. Introducción

En las últimas dos décadas el interés por el estudio de la conectividad de la sociedad y otros sistemas tanto naturales como artificiales ha cobrado un extraordinario impulso. La aparición y el crecimiento de Internet, la propagación de la crisis entre los países, de enfermedades entre individuos o la difusión de noticias e ideas son algunos ejemplos que suscitan el interés del estudio de cómo determinadas conexiones de las redes intervienen en, o favorecen, estos fenómenos. Aunque este estudio ha adquirido especial importancia en las últimas dos décadas debido a los recientes avances tecnológicos, numerosos investigadores de diferentes disciplinas comenzaron su estudio mucho antes. Los *sociogramas* de Levy Moreno, el experimento del *mundo pequeño* de Stanley Milgram, también conocido como los *seis grados de separación* (1967) o el trabajo de los físicos Watts y Strogatz (1998) son algunos ejemplos de importantes contribuciones que han permitido alcanzar la visión que tenemos ahora del estudio de redes sociales.

### 1.1 Motivación

Recientemente, con la emergencia, proliferación y el éxito de las aplicaciones sociales como blogs (p. ej. Blogger, WordPress), redes sociales online (p. ej. Facebook, MySpace) o microblogs (p. ej. Twitter) entre otras, se ha renovado y redefinido el interés por el estudio de la influencia social y cómo ésta se propaga por las redes. La propagación de influencia es un campo de investigación abierto de por sí, en el que hay un amplio espacio por explorar para llegar a comprender cómo se producen estos fenómenos. La asunción principal es que cuando las personas ven a sus conocidos realizar una acción ellos pueden decidir realizar una acción igual o parecida. Sin embargo, el motivo por el que pueden haber decidido realizarla es porque sea muy popular, por haberlo oído fuera de la red social o porque en verdad han sido influidos por ver a sus contactos desempeñar esta acción. Discernir si existe este tipo de influencia es todo un reto en la actualidad y motivo de numerosos estudios.

La motivación del presente trabajo se enmarca dentro de en avanzar en la comprensión de los fenómenos de propagación de influencia en las redes sociales, atendiendo a aspectos como la velocidad de transmisión, las características de los estados estacionarios, y en general el desarrollo de la dinámica de los sistemas a medida que se propaga una señal. Veremos cómo, p. ej., el análisis de las propiedades estructurales de las redes permite ver qué factores condicionan la propagación de la información además de su velocidad de transmisión (Doerr, Fouz, & Friedrich, 2012) o el poder de la presencia de los enlaces fuertes y débiles en el grafo (Kleinberg & Easley, 2010).

Existe además una relación potencial, apenas estudiada, entre la propagación de influencia en una red social y el efecto de los sistemas de recomendación. La introducción de los recomendadores en las redes sociales tiene su repercusión en cómo los usuarios se ven influidos unos por otros y, potencialmente, por las propuestas que envía el propio recomendador. En este sentido contemplamos los sistemas de recomendación en este trabajo bajo la perspectiva de un agente que puede modificar la forma en la que se propague la información y el modo en que los usuarios descubran opciones de su interés.

Una vez abordados los temas anteriores y en base al estudio realizado, se propone y explora en este trabajo un modelo de propagación de influencia desde un enfoque no estudiado previamente. Se trata de un modelo de agentes basado en el modelo propuesto por Lallouache *et al.* (2010) para simular las dinámicas colectivas de opinión. El nuevo enfoque permite adaptar el escenario a múltiples opciones de opinión (preferencias multivaluadas), propio de los sistemas de recomendación, que operan en espacios donde se considera una relación de preferencia entre usuarios y diferentes artículos (ítems) entre los que los usuarios optan según sus gustos, pasando así de una única opinión al estudio de la dinámica de múltiples opiniones. El escenario es propio en general de cualquier situación en la que las personas se forman una preferencia por diferentes alternativas (opciones políticas, de opinión, de consumo, etc.) a través de su propia experiencia y de la influencia de otras personas e informaciones con las que se relaciona. El modelo se ha implementado en un banco de prueba que permite su simulación, además de su visualización, facilitando su estudio y comprensión. Mediante una herramienta desarrollada al efecto, hemos estudiado la propagación de opiniones en función de ciertos parámetros, de las características de la red y las condiciones iniciales, mediante el desarrollo de un modelo y de un simulador, permitiendo la experimentación comparativa del efecto producido por los parámetros seleccionados, así como la visualización del proceso de formación de opinión.

En los resultados obtenidos se observan diferentes situaciones en la formación de opinión del conjunto de individuos del grafo. Por un lado, existen estados en los que la preferencia de los individuos es homogénea, es decir, tras varias ejecuciones todas las personas del grafo prefieren el mismo ítem. Sin embargo, en función de los parámetros esta preferencia puede variar, siendo en ocasiones de unos pocos ítems y en otras permaneciendo todos, no existiendo presencia de influencia alguna. Estos efectos han sido comprobados tanto en redes simuladas generadas con diferentes modelos aleatorios y parámetros como en redes sociales reales. Todos estos efectos pueden constataarse tanto de forma visual como en las métricas aplicadas.

Adicionalmente, se ha probado el efecto de la introducción de un agente recomendador en el sistema. Puesto que no existe información adicional de los usuarios que permita aplicar técnicas de recomendación como filtrado colaborativo, se ha seleccionado un recomendador basado en popularidad, es decir, recomendando el ítem preferido por una mayoría simple de individuos en cada momento. Como veremos, este agente provoca la ruptura de la formación de clústers de opinión en redes reales, que se caracterizan por la presencia de comunidades altamente cohesionadas en conectividad, además de incrementar la aparición de escenarios en los que los individuos tienen las mismas preferencias.

## **1.2 Objetivos**

El presente trabajo tiene como objetivos principales,

- Obtener una visión global avanzada de los fenómenos de propagación en redes sociales, una perspectiva en amplitud de la tipología, los diferentes paradigmas y modelos definidos en esta área para describir la fenomenología, y los problemas abordados en relación con los mismos.

- Desarrollar un modelo formal de propagación de influencia e implementar este modelo en un marco de prueba que permita su simulación. Observar el comportamiento del sistema en función de los parámetros introducidos.
- Estudiar el efecto de la introducción de un 'agente' recomendador en el proceso de formación de opinión.

Las secciones en las que se estructura el trabajo son las siguientes: en primer lugar, se realiza una introducción al estudio de redes sociales repasando brevemente el contexto histórico y las métricas y modelos de generación de grafos más utilizados. A continuación, profundizaremos en fenómenos que tienen lugar en las redes sociales, focalizándonos en procesos de propagación e influencia. Posteriormente, se detalla el experimento llevado a cabo, describiendo la metodología, la definición del modelo, las pruebas realizadas y los resultados observados. Se concluye el trabajo con unas conclusiones de lo realizado, así como direcciones de trabajo futuro.



## 2. Estudio de las redes

El estudio científico de las redes, como las redes de ordenadores, las redes biológicas o sociales, es un campo interdisciplinar que combina ideas de matemáticas, física, biología, informática y muchas otras áreas. Hay muchos sistemas de interés para los científicos que se componen de partes individuales unidas de alguna forma: p.ej. Internet (colección de ordenadores unidas por conexiones de datos), las sociedades humanas (conjunto de personas unidas por interacciones sociales), la red de teléfonos, etc. Todos estos ejemplos, aunque distintos, tienen características en común: en todos los casos hablamos de objetos (p.ej. ordenadores o individuos) que se encuentran conectados formando una red.

### 2.1 Contexto histórico

La fecha en la cual los investigadores comenzaron a utilizar el concepto de red social es difícil de determinar (Scott & Carrington, 2011).

En antropología y sociología, las redes sociales han sido motivo de estudio en diferentes campos desde aproximadamente el segundo cuarto del siglo pasado. Aunque son muchos los investigadores, su fundación se atribuye a Jacob Levy Moreno, quien desarrolló la denominada *sociometría* en la década de los 30, un método cuantitativo para medir las relaciones sociales. Una de las innovaciones de Moreno fue la invención del sociograma, un diagrama en el que los actores se representan como puntos y las relaciones existentes entre ellos como líneas, mostrando de forma gráfica las distintas relaciones entre los sujetos que forman un grupo. Estos diagramas, a excepción del nombre, son lo que actualmente conocemos como redes sociales.

En el desarrollo posterior del análisis de redes sociales cabe mencionar los estudios de parentesco de Elizabeth Bott en Inglaterra entre los años 1950. Bott (1955) estudió la vida de 20 de familias británicas, especialmente las relaciones de parentesco, a través del diseño de redes empleando ideas de conectividad y densidad. Hacia mediados de esa misma década un grupo de antropólogos de la Universidad de Manchester realizaron asimismo estudios de urbanización, investigando redes comunitarias en el sur de África, India y el Reino Unido.

Una de las contribuciones más conocidas en el trabajo temprano en este campo fue la realizada por el psicólogo Stanley Milgram en los años 60 con su experimento del *mundo pequeño*, también conocido como los *seis grados de separación* (1967). En este experimento Milgram trataba de cuantificar la distancia entre dos personas aleatorias en la red social formada por todas las personas que se conocen entre sí. Para llevar a cabo el experimento, envió 96 paquetes a diferentes personas (escogidas de forma aleatoria) de la ciudad de Omaha, Nebraska. El paquete contenía un pasaporte de un amigo suyo de Boston, Massachusetts, además de una serie de datos como su nombre, dirección y ocupación. En las instrucciones se pedía a los participantes que, para hacer llegar el paquete a su destino, no podían enviarlo directamente al destinatario, y sólo debían entregarlo a un conocido suyo del cual conocieran su nombre y pensarán que haría avanzar el paquete hacia su destino. Esta persona debía seguir las mismas instrucciones. Al final del experimento llegaron 18 paquetes a su destino y descubrió que la longitud media del camino entre las personas de Omaha hasta su

amigo era de 5.9 pasos. De este experimento surgió la hipótesis de los *seis grados de separación*.

Los años 70 marcan un gran momento de crecimiento del área, con el nacimiento de la INSNA (*International Network for Social Network Analysis*) fundada por Barry Wellman en 1977, de dos revistas, *Connections* y *Social Networks* y una gran producción bibliográfica. En esta década también se producen grandes avances en el campo teórico, metodológico y conceptual de las redes. Destacan los trabajos de Bearden *et al.* (1975) los cuales, basándose en la idea de centralidad, exploraron el poder y la influencia de los bancos en el mundo empresarial americano unido.

Desde finales de la década de los 70, ha tenido lugar un enorme incremento en las contribuciones técnicas a la metodología de las redes sociales y a sus aplicaciones (Burt, 1982).

El desarrollo más sorprendente en los últimos años ha sido el crecimiento del interés de los físicos en aplicar ideas de redes a fenómenos sociales. Este interés se manifiesta por primera vez en el trabajo de Watts y Strogatz (1998) basado en el experimento de Milgram de los *seis grados de separación* (1967) y en la literatura de redes aleatorias que se ha desarrollado a partir de él. Barabási y Watts sugirieron que sus ideas podrían aplicarse al mundo social, ignoraron por completo todo el trabajo realizado hasta ese momento por los sociólogos, antropólogos y científicos tanto políticos como económicos.

En la actualidad Watts ha roto con esta división, moviéndose a la sociología y sus ideas han sido debatidas por los sociólogos. Barabási, en cambio, continúa ignorando todo el trabajo realizado previamente en este campo. Sin embargo, el trabajo realizado por los físicos ha permitido poner el punto de mira en aspectos que no habían sido profundizados previamente por los científicos sociales.

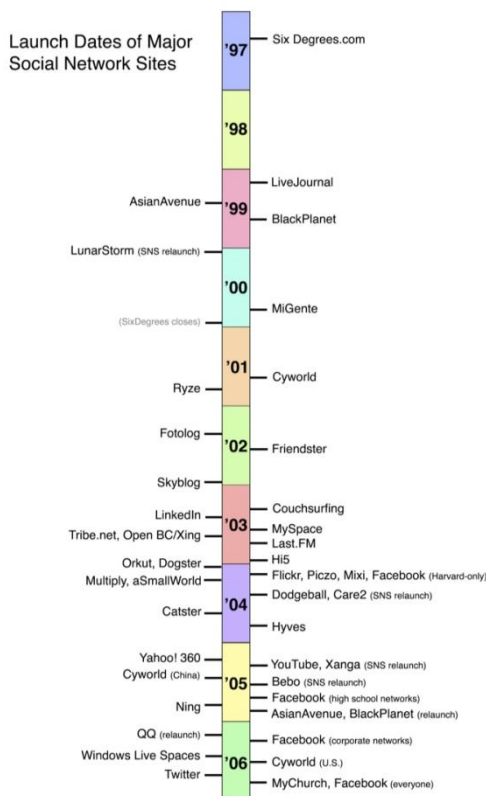


Figura 1 - Hitos en el desarrollo de las redes sociales online entre 1997 y 2006 (Boyd & Ellison, 2008)

En los últimos años, el desarrollo de internet, la generalización de la disponibilidad de ordenadores y de las bases de datos online, han permitido la construcción de numerosas redes de diferentes tipos y escalas, en órdenes de magnitud hasta entonces desconocidos. Además, recientemente han surgido las redes sociales online como Facebook o LinkedIn, que permiten explotar las redes de contactos entre individuos y que han revolucionado el concepto que teníamos hasta ahora de las redes sociales.

En la imagen se muestran las fechas de lanzamiento de las principales redes sociales online desde 1997 hasta el 2006. Puede verse cómo en los últimos años se ha producido un boom en el lanzamiento de estos sitios web.

## 2.2 Estado actual

El análisis de redes sociales se ha convertido en un área de estudio interdisciplinar, con su propia asociación profesional, conferencia anual y múltiples revistas. El INSNA ha crecido desde los 175 socios fundadores hasta más de 1300 en febrero de 2011 y su conferencia anual (*International Sunbelt Social Network Conference*) atrae cada año a más de 500 personas.

Tres revistas publican, exclusivamente, investigaciones en redes sociales: *Social Networks*, *Connections* y *Journal of Social Structure* (publicada online). Además, en la mayoría de las revistas científicas (*American Journal of Sociology*, *Social Forbes* o *American Sociological Review* entre otras) aplican una perspectiva de redes sociales en las investigaciones.

Todos los avances vistos en la sección 2.1, han permitido a los investigadores acceder a gran cantidad de datos para analizar en formato digital que antiguamente no estaban disponibles o sólo se podían adquirir mediante costosos experimentos como el de Milgram (1967). La disponibilidad de datos abre un amplio conjunto de posibilidades para estudiar la estructura de las redes sociales, el comportamiento de los individuos, la presencia de comunidades latentes, la eficiencia de la recomendación viral o la detección de los usuarios con mayor poder de influencia, entre otras. Todas estas aplicaciones serán abordadas a lo largo de los siguientes puntos del trabajo.

El núcleo del análisis de las redes sociales es una rama de las matemáticas llamada *la teoría de grafos* que ha dado base a la representación y fundamentación básica de las redes. En esta teoría, el conjunto de objetos equivale a los *nodos* y las relaciones se muestran como *aristas*. Formalmente, un grafo es un par  $G = (V, A)$  formado por un conjunto de nodos  $V$  y un conjunto de arcos  $A$ .

Estos enlaces, dependiendo del contexto en el que nos encontremos, tendrán un significado diferente como se muestra en la Tabla 1.

Network	Vertex	Edge
Internet	Computer or router	Cable or wireless data connection
World Wide Web	Web page	Hyperlink
Citation network	Article, patent or legal case	Citation
Power grid	Generating station or substation	Transmission line
Friendship network	Person	Friendship
Metabolic network	Metabolite	Metabolic reaction
Neural network	Neuron	Synapse
Food web	Species	Predation

Tabla 1 -Vértices y aristas en redes (Newman M. , 2010)

Actualmente, existen numerosos tipos de redes sociales además de las redes online explícitas, tales como las redes de interacción (en comunicación por teléfono, mail, etc.), redes de colaboración (compañeros de trabajo, actores, etc.), redes de cita bibliográficas, etc. Todos estos tipos de redes muestran relaciones que pueden ser simétricas o asimétricas, o dicho en términos de la teoría de grafos, *dirigidos* o *no dirigidos* (Figura 2). Ejemplo de redes simétricas son las amistades de la red social Facebook o los contactos en LinkedIn. En Twitter por el contrario las relaciones son asimétricas: un usuario  $A$  puede seguir a  $B$  pero puede que  $B$  no siga al usuario  $A$ .

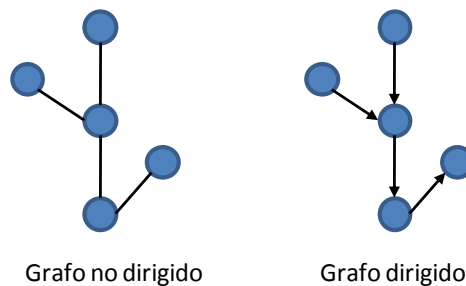


Figura 2 - Grafo dirigido vs no dirigido

Existen diferentes aspectos que son estudiados en el análisis de las redes los cuales se irán abordando en los siguientes puntos.

## 2.3 Propiedades estructurales de las redes

El propósito de estudiar las redes es comprender mejor el comportamiento de los sistemas que representan. La topología de la red es uno de los aspectos determinantes a la hora de definir el comportamiento de las redes ante diversos fenómenos, por lo que el análisis de la misma nos ayuda en su comprensión: por ejemplo, la forma en la que se conectan los ordenadores en una red afecta a la eficiencia con la que se transporta la información, al igual

que la forma en la que los usuarios se comunican en una red social interviene en cómo las personas se forman las opiniones o son informadas de noticias.

En el estudio las características de los nodos, existen diferentes métricas que permiten determinar su importancia en la red. Algunas de las métricas más utilizadas son las relacionadas con el estudio del grado, con la centralidad y la cohesión.

## Métricas locales

### Grado

El grado de un nodo  $u$ , representado como  $g(u)$ , hace referencia al número de aristas a las que está conectado. En grafos dirigidos surge la distinción entre *indegree* (grado entrante, es decir número de aristas que inciden en el nodo) y de *outdegree* (número de aristas que tienen su origen en el nodo). Esta distinción se muestra en la Figura 3,

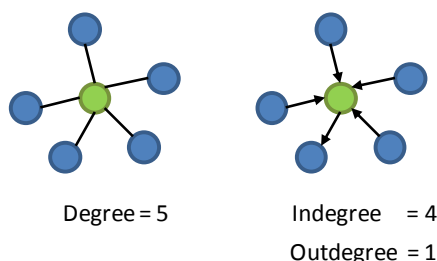


Figura 3 - Ejemplo del grado de un nodo

En grado, aunque es una métrica muy sencilla, proporciona una idea de la importancia del nodo, y resulta ya de por sí muy significativa respecto a su papel en la red.

### Coefficiente de clustering local

El principio del cierre triádico por el cual si dos usuarios comparten un amigo la probabilidad de compartir un enlace se incrementa (“los amigos de mis amigos son mis amigos”), ha motivado la definición de métricas como el coeficiente de clustering.

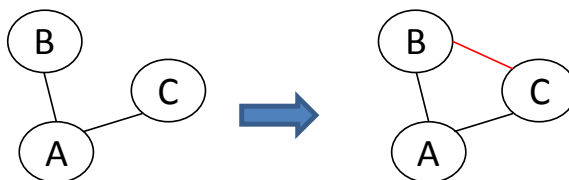


Figura 4 - Cierre triádico

El coeficiente de clustering de un nodo  $u$ , representado como  $C(u)$ , mide la probabilidad de que dos vecinos del nodo tomados de forma aleatoria sean vecinos. Es una forma de medir cómo de completo es el grafo en torno al nodo.

$$C(u) = \frac{n^{\circ} \text{ conexiones entre vecinos de } u}{n^{\circ} \text{ conexiones posibles entre vecinos de } u} \in [0,1]$$

Donde el número de conexiones posibles entre vecinos de  $u$  viene dado por:

$$\begin{cases} g(u)(g(u) - 1) & \text{si } G \text{ es dirigido} \\ \frac{g(u)(g(u) - 1)}{2} & \text{si } G \text{ no es dirigido} \end{cases}$$

## Métricas globales

### Centralidad

Dentro del concepto de “centralidad” de los nodos se consideran las nociones de *betweenness* y *closeness*.

- *Betweenness*: mide en qué medida un nodo es punto de paso de muchos caminos entre otros usuarios de la red. Para ello, se basa en el número de caminos de distancia mínima (CDM) que pasan por él: cuántos más caminos pasen, mayor valor de *betweenness*.

$$B(u) = \frac{2}{(n-1)(n-2)} \sum_{\substack{v,w \neq u \\ v < w}} \frac{ns_{v,w}(u)}{ns_{v,w}}$$

Donde,

$ns_{v,w}$  es el número de CDM entre los nodos  $v$  y  $w$ .

$ns_{v,w}(u)$  es el número de CDM entre los nodos  $v$  y  $w$  que pasan por  $u$ .

Aquellos nodos con un valor alto de esta métrica juegan un papel estratégico en el paso de información y de potencial influencia en el grafo por su posición de intermediarios en la comunicación entre usuarios.

- *Closeness*: mide en qué medida el nodo se encuentra en una posición cercana, en promedio, a todos los demás nodos del grafo.

$$C(u) = \frac{n-1}{\sum_{v \neq u} d(u,v)}$$

Al igual que la métrica anterior, refleja una posición de influencia puesto que permite llegar rápidamente a todos los nodos.

Closeness y *betweenness* son distintas formas de identificar nodos que ocupan una posición central en la red social, y no tienen por qué correlacionar entre sí. Mientras que *closeness* refleja una centralidad que podríamos denominar “geométrica” o de equidistancia, *betweenness* capta una propiedad relacionada con flujos de red y el enlace entre comunidades.

### Coefficiente de clustering global

El coeficiente de clustering global, representado por  $C(G)$ , es la probabilidad de que dos nodos del grafo con un amigo en común tomados al azar sean vecinos. Refleja la unión de entornos de la red y su valor dependerá de cómo se han formado los enlaces de la red. Si la formación

se ha producido de forma aleatoria el coeficiente será bajo. En cambio, su valor aumentará si los enlaces se forman por amistad o similitud.

Para calcularlo,

$$C(G) = \frac{n^{\circ} \text{ caminos cerrados de long } 2}{n^{\circ} \text{ de caminos de long } 2} \in [0,1]$$

### **Asortatividad**

La assortatividad representa en qué medida los usuarios se relacionan con usuarios similares (homofilia) o diferentes (heterofilia). Existen diferentes tipos: assortatividad por similitud de tipo, escalar o de grado.

La assortatividad por similitud de grado se encarga de medir la correlación entre el grado de un nodo y el de sus vecinos. Sirve para ver en qué medida los usuarios se relacionan con gente similar a ellos o distinta. Por ejemplo, en el caso de Twitter no se suele dar ya que puedes tener mucha gente que te siga y que a ellos no les siga tanta gente. Pero en el caso de en redes como Facebook si se da: sueles conectarte a gente con un número de contactos similares.

### **Conectividad del grafo**

Dado un grafo es natural preguntarse si cada par de nodos se encuentra conectado mediante un camino, es decir, si el grafo está conectado. En casos como una red de transporte es esperable que todos los elementos estén conectados. Sin embargo, en una red social puede que dos personas no estén conectadas por ningún camino. En este caso, el grafo se encuentra desconectado, formado por diferentes componentes. El término componente hace referencia a un subconjunto de nodos en el cual cada par de nodos se encuentra conectado y el subconjunto no es parte de otro conjunto mayor de nodos.

Cuando una de las componentes conexas abarca una alta proporción de los nodos se habla de *componente gigante*. Esta componente se muestra marcada en rojo en la Figura 5.

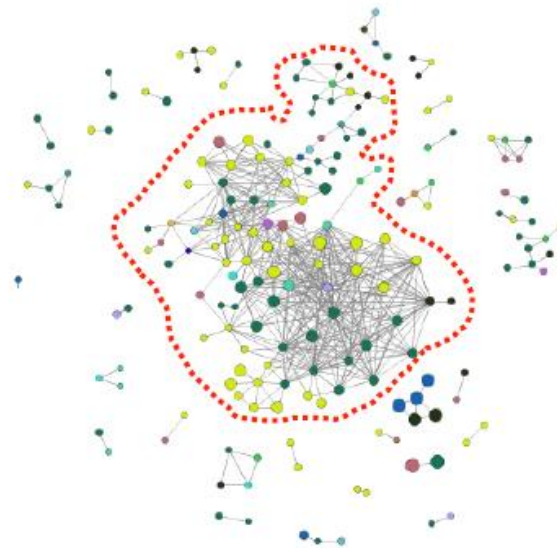


Figura 5 - Componente gigante en una red

## 2.4 Modelos de formación de redes

Otra área fundamental del estudio de las redes sociales ha estado en el análisis y modelización de los procesos por los que se forman las redes, las propiedades que resultan de los mismos, y el poder de explicación y predicción que tienen en relación con las redes reales. Los modelos de formación de redes nos permiten reproducir o simular redes que se parecen a la red con la que queremos trabajar. Los modelos ayudan además a explicar y predecir respuestas de las redes a determinados fenómenos sobre ellas. Los modelos son una herramienta con la que podemos simular (teórica o computacionalmente) procesos de red sobre una base analítica, que facilita expresar de forma directa determinados resultados sin necesidad de computarlos sobre los datos reales. Sabiendo, por ejemplo, que una red tiene una determinada distribución de grado, podemos estimar qué efecto tendrá sobre ella un proceso de propagación de información.

A continuación se explican algunos de los modelos importantes propuestos y estudiados en el área, de los que haremos uso en el trabajo experimental del presente TFM.

### Modelo Erdős-Rényi

Este modelo, realizado por los matemáticos Paul Erdős y Alfréd Rényi (1959), asume que la red es no dirigida y que los enlaces se unen de forma aleatoria. El parámetro más importante es  $p$ , que representa la probabilidad de que dos nodos compartan una arista.

$$\text{Dados } u, v \in V, p((u, v) \in A) = p$$

Es el modelo más sencillo y aleatorio. Como características:

- Tiende a generar una componente gigante muy pronto (con  $avg_u g(u) \sim 3$ )
- Tiende a generar un coeficiente de clustering muy bajo lo cual no es un fenómeno muy común en las redes reales.
- El grado de los nodos sigue una distribución binomial.



$$p(g(u) = k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

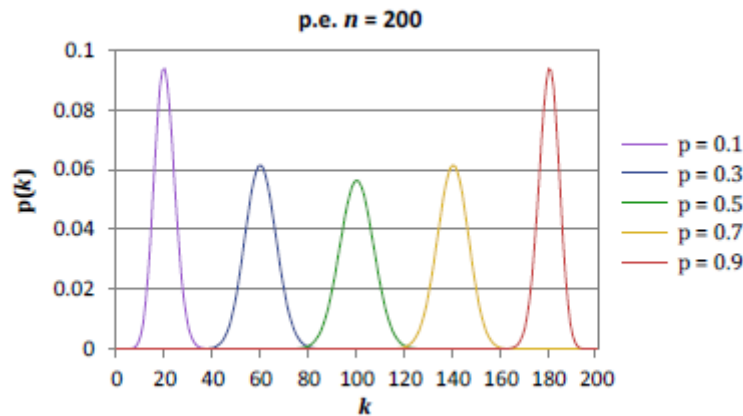


Figura 6 - Distribución binomial del grado de los nodos (ER)

- Debido a esta distribución, en estos modelos hay ausencia de comunidades (agrupamientos) y de hubs (nodos con grado muy superior a la media).

### Enlace preferente: Barabási - Albert

El modelo de Barabási-Albert, propuesto por Albert-László Barabási y Réka Albert (1999) es un modelo de enlace preferente o también conocidos como “rich gets richer”.

En el modelo de enlace preferente los nodos van apareciendo uno a uno y seleccionan  $m$  nodos a los que conectarse. Esta elección no es uniforme sino que la probabilidad de que un nodo  $u$  se enlace a un nodo  $v$  es proporcional al grado de  $v$ . Uno de los efectos es que los enlaces más antiguos acumularán más enlaces.

Este comportamiento se observa en las redes reales ya que, por ejemplo, en el caso de Facebook, existen contactos que acumulan más enlaces que el resto (son más populares). También en Twitter donde hay un número reducido de personas/grupos que son seguidos por la mayoría mientras que el resto tienen muchos menos seguidores.

La Figura 7 muestra las diferencias entre la creación de un grafo aleatorio (a la izquierda) y otro con un modelo de enlace preferente con  $m = 2$  (a la derecha).

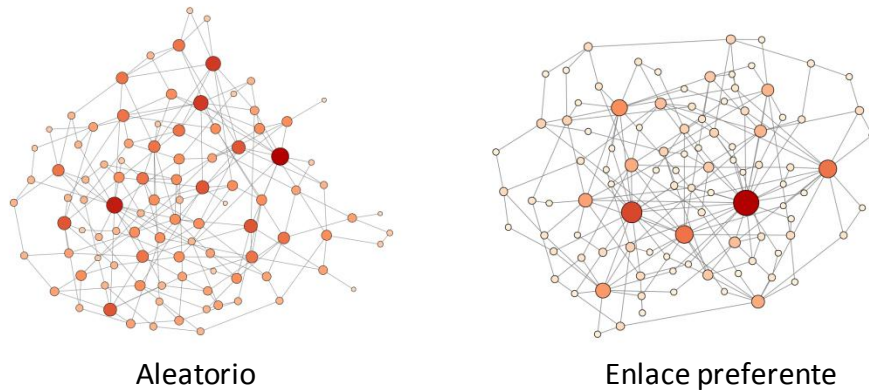


Figura 7 - Modelo aleatorio vs enlace preferente

Como se aprecia en la Figura 7, en el grafo generado mediante enlace preferente existen nodos con más grado que otros. En el grafo aleatorio estas diferencias no son tan destacables.

Los modelos de enlace preferente tienen una distribución de grado power law. Es decir, existen unos pocos nodos que acumulan todos los enlaces, frente al resto que tiene muchos menos enlaces.

Este fenómeno se produce de forma repetida en fenómenos naturales (Newman M. E., 2005): los enlaces sociales y de la web, las consultas realizadas o los productos adquiridos son algunos ejemplos, en los que se una “short head” de unos pocos elementos acumulan una gran parte de la masa de la distribución, mientras que el resto de elementos forman una “cola larga” donde la gran mayoría de elementos tiene una muy baja probabilidad. En sentido estricto una distribución power law es,

$$p(k) = Ck^{-\alpha}, \text{ con } \alpha > 0$$

Donde,

- $p(k)$  es la probabilidad de observar un nodo de grado  $k$ .
- $C$  es una constante de normalización.
- $\alpha$  es el exponente de la power law. Los valores típicos en la naturaleza son  $2 \leq \alpha \leq 3$ .

La red generada por el modelo de Barabási-Albert es una red power law, también denominada “libre de escala”, con exponente  $\alpha = 3$ .

$$p(k) = Ck^{-3}$$

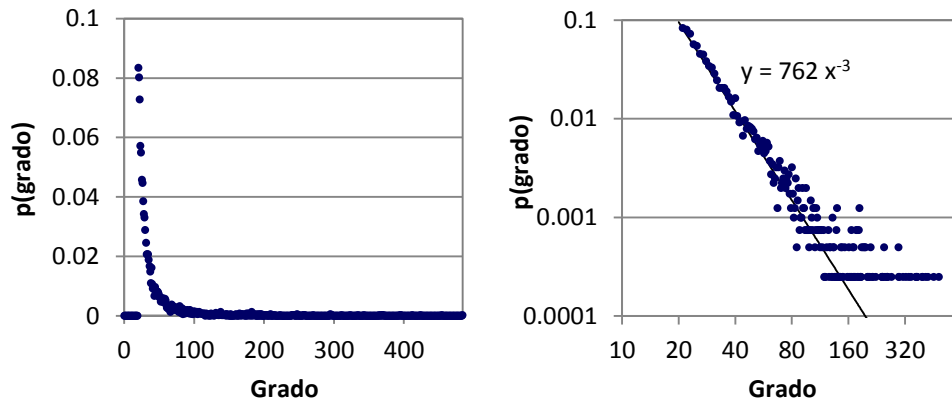


Figura 8 - Distribución power law del grado en un grafo Barabási-Albert de 4000 nodos con grado promedio 40

Aunque, como hemos visto anteriormente, este proceso está presente en fenómenos naturales, existen algunas diferencias con las redes reales. Estos modelos generan grafos con un coeficiente de clustering menor y su grado crece indefinidamente con la antigüedad del nodo, características que marcan diferencias con las redes reales.

### **3. Procesos en las redes**

En la sección anterior hemos visto cómo el estudio de las propiedades estructurales ayuda en la comprensión de la red. Sin embargo, hablar de conectividad no sólo implica ver quién está conectado con quién, sino también analizar el comportamiento de los usuarios y las dinámicas globales que emergen a partir de éstos; las acciones de un usuario tienen consecuencias en el resto de individuos de la red que pueden ser estudiadas de forma agregada. Las redes también nos pueden ayudar a predecir y a extraer conclusiones de cómo se va a comportar el sistema. Los desarrollos teórico-prácticos en esta área incluyen líneas como el estudio de epidemias, los sistemas dinámicos en redes y los procesos de difusión.

En los siguientes apartados se explica uno de los procesos más sencillos: la percolación. También se hablará de procesos de difusión en las redes y de un caso concreto de este tipo de procesos: los modelos de epidemias.

#### **3.1 Percolación**

El término de ‘percolación’ se ha tomado del campo de la física para referirse en el contexto de las redes sociales a procesos que se describen por la eliminación sucesiva de aristas o vértices (junto con las aristas que los unen) de un grafo con el propósito de estudiar el comportamiento emergente. Este concepto se manejó hace unos cincuenta años en el campo de la física referido al paso de un líquido a través de un medio poroso desordenado, es decir, con canales bloqueados al azar –de ahí la analogía (en la naturaleza del fenómeno, la formulación de los problemas y el desarrollo de soluciones) con la supresión de conexiones en las redes sociales.

Existen diferentes formas en las que se pueden ir seleccionando los vértices de un grafo en un proceso de percolación. El caso más sencillo es seleccionar de forma aleatoria aquellos vértices que no forman parte de la funcionalidad del grafo. Otra forma de seleccionarlos es mediante el grado de los nodos (por ejemplo, eliminándolos de mayor a menor grado).

Para observar el efecto nos basamos en el siguiente ejemplo,

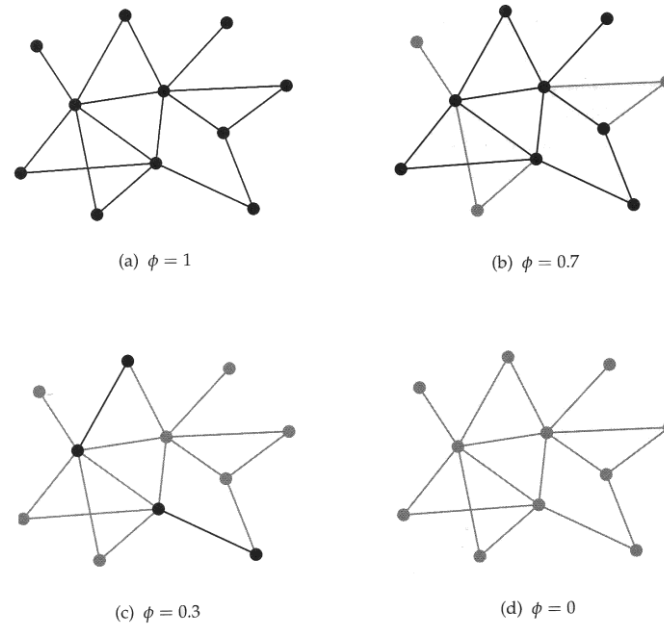


Figura 9 - Ejemplo de percolación (Newman M. , 2010)

La Figura 9 representa el proceso de percolación en una red para diferentes valores de  $\phi$  es decir, de la probabilidad de que un vértice esté presente o funcionando ( $\phi=1$  indica que todos los vértices están presentes y  $\phi=0$  indica que todos los vértices han sido eliminados). Los vértices marcados en gris indican que han sido eliminados, junto con sus aristas, y aquellos representados en color negro indican que los vértices están presentes. Como se observa en la imagen (a) todos los vértices se encuentran unidos en una componente gigante. Según va disminuyendo el valor de  $\phi$ , es decir, según se eliminan vértices y sus aristas, llega un punto en el que la componente se rompe, dejando sólo componentes pequeñas. La formación o ruptura de la componente gigante se denomina *percolation transition* y el punto en el cual se produce *percolation threshold*.

El estudio de la percolación se ha enfocado a temas tales como su impacto en la pervivencia o ruptura de una componente conexa gigante, o en dinámicas de propagación (frenándola a partir de un punto). Un ejemplo del efecto de percolación es la vacunación de un individuo inmunizándolo y, por lo tanto, evitando que transmita una enfermedad. Al hacerlo, el individuo deja de ser transmisor de la enfermedad aunque continúe presente en la red, por lo que equivale a extraer el vértice que representa. Este efecto tiene consecuencias no sólo en el individuo o grupo de individuos a los que se vacuna, sino que afecta a la red entera. Por este motivo, vacunar a una parte de la población puede suponer que se consiga la inmunidad de la red.

En la actualidad la forma normal de expresión de las teorías de la percolación, aunque con diferencias en la terminología, es la teoría de grafos. Los vértices se llaman sitios (*sites*) y los arcos y enlaces o ligaduras (*bonds*); cuando se obtiene un grafo seleccionando nodos se habla de percolación de sitio (*site percolation*) y cuando se seleccionan los arcos, percolación de enlace (*bond percolation*).

Es importante elegir qué modelo captura mejor la naturaleza del problema puesto que difiere bastante en casos como, por ejemplo, las predicciones sobre la difusión de una epidemia. En el caso de un virus como el Ébola se puede asumir que todo el mundo es susceptible de contraer la enfermedad y se quiere poner el foco en la medida en que cada quien puede infectar a otros (modelo de percolación de ligadura). Sin embargo, los virus de ordenador como el Melissa, pasaron generalmente entre una computadora susceptible y otra (todas las ligaduras están efectivamente abiertas) pero no todas las computadoras son susceptibles (modelo de percolación de sitio).

### 3.2 Sistemas dinámicos en redes

Los sistemas dinámicos hacen referencia a un sistema cuyo estado, representado por un conjunto de variables, cambia a lo largo del tiempo en función de unas reglas o ecuaciones. La propagación de enfermedades, de noticias o rumores son un caso particular de sistema dinámico. Otros ejemplos son el movimiento de los datos en Internet, de la electricidad en la red, del tráfico en las calles y numerosos sistemas de gran interés para estudiar.

#### Epidemias en redes

Uno de los aspectos estudiados en las redes sociales es su papel en el contagio de enfermedades. Aunque para entender el proceso de contagio de las enfermedades habría que considerar ciertos aspectos biológicos, estudiar la forma en la que los individuos se van contagiando a través de las redes sociales ayuda a predecir y comprender la evolución y el resultado de las epidemias. Para ello, existen diferentes modelos de contagio.

El *modelo SI* es la versión más sencilla y se basa en la existencia de dos posibles estados: *susceptible (S)* e *infectado (I)*. Si un individuo es susceptible significa que no tiene la enfermedad pero que puede contraerla si permanece en contacto con alguien que la tenga. Si está infectado significa que tiene la enfermedad y que puede contagiarla si entra en contacto con una persona susceptible. Este modelo, aunque simple, permite ver qué ocurre a nivel de la población.

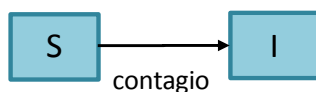


Figura 10 - Transiciones permitidas en el modelo SI

El *modelo SIR* refleja un estado más realista para cierto tipo de enfermedades. Añade un nuevo estado, *recuperado (R)*, en el que el sistema inmunológico del individuo ha combatido la enfermedad. Además se vuelve inmune, con lo que no podrá volver a coger la enfermedad.



Figura 11 - Transiciones permitidas en el modelo SIR

El *modelo SIS* extiende del modelo SI y refleja el comportamiento que tienen ciertas enfermedades en las que no existe inmunidad como en el modelo SIR, pero si puede haber

recuperación. En este modelo se permite la *reinfección*, es decir, una vez recuperado puedes volver a contraer la enfermedad.

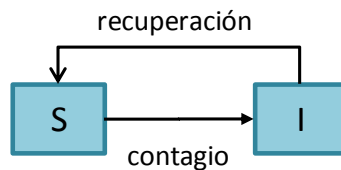


Figura 12 - Transiciones permitidas en el modelo SIS

Por último, el *modelo SIRS* permite también que los individuos adquieran inmunidad a las enfermedades como ocurre en el modelo *SIR*. Sin embargo, esta inmunidad es temporal y al cabo de un cierto tiempo el individuo puede volver a infectarse.

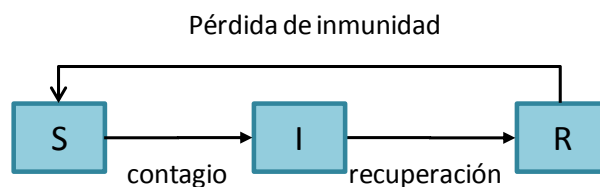


Figura 13 - Transiciones permitidas en el modelo SIRS

### Propagación en redes

Uno de los focos de estudio y en el que centramos el trabajo es el análisis de cómo se propagan la información y las opiniones en las redes sociales y, concretamente, cómo cada individuo es influido por sus vecinos en la red social.

Las redes sociales como Facebook y Twitter, debido a su reciente auge y el número masivo de usuarios que la utilizan, juegan un papel muy importante en la difusión de información. Bakshy *et al.* (2012) analizan el papel que juegan las redes sociales en la difusión de información mediante un experimento realizado con Facebook. En el experimento, los usuarios son divididos en dos grupos: aquellos que reciben las noticias por la red social y aquellos a los que se les ocultan estos enlaces y, por lo tanto, han de recibirlo de forma externa. Observan que aquellos que ven compartir los enlaces a sus amigos es más probable que compartan la misma información (7,37 veces) y que la compartan antes que aquellos que no disponen de esta información por las redes sociales.

El efecto de propagación no sólo se observa en información y opinión, también tiene lugar en otros tipos de estados que afectan a las personas en diferentes ámbitos. Por ejemplo, se pueden propagar hábitos o decisiones entre otras.

Como hemos visto en el apartado 3.2, también se pueden propagar virus y enfermedades donde las conexiones entre los individuos de una red pueden verse como el medio sobre el que se propaga la enfermedad. De esta forma, dos individuos que respiran el mismo aire pueden contraer una enfermedad infecciosa respiratoria o mediante el contacto físico contraer parásitos.

Si comparamos la propagación de ideas con los modelos de contagio vemos que existen numerosos aspectos en común: tanto las ideas como las enfermedades se pueden transmitir de individuo a individuo a través de las redes que conectan personas. El motivo de tratarlas de forma separada reside en la forma en la que los individuos adoptan la idea/enfermedad. Cuando hablamos de modelos de contagio, los individuos no eligen ser infectados. Se trata de un proceso que no es observable y el contagio se modela de forma aleatoria (si estás conectado, existe una probabilidad dada de infectarte). En cambio, al hablar de difusión de ideas, existe un proceso de decisión en el que el usuario puede o no adoptar la idea. En muchos otros aspectos es posible no obstante establecer paralelismos, y desarrollar teorías y métodos comunes válidos para ambos tipos de escenario.

En la siguiente sección profundizaremos en esta idea, abordando los distintos modelos de propagación existentes y cómo los usuarios se ven influidos a la hora de tomar sus propias decisiones.



## 4. Propagación e influencia en las redes

En este apartado revisamos los modelos, teorías y estudios principales sobre cómo los comportamientos, ideas, prácticas y opiniones se propagan de persona en persona a través de la red, cómo las personas influyen a sus amigos para adoptar nuevas ideas.

Para comprender este proceso es necesario entender cómo fluye la información a través de las redes, el papel importante que juegan determinados nodos en este proceso o cómo las características estructurales marcan su evolución son aspectos determinantes.

### 4.1 Aspectos determinantes

#### Propagación y topología

Las propiedades estructurales de las redes son uno de los factores que condicionan la propagación de la información y su velocidad de transmisión.

Así por ejemplo, en una serie de experimentos y análisis matemáticos, *Doerr et al.* (2012) observan que la velocidad a la que se propaga un rumor es diferente en función de la topología del grafo. Para la prueba utilizan tanto redes sociales reales como topologías clásicas de grafos y llegan a la conclusión de que aquellos nodos con grado mayor tardan más en propagar el rumor que los nodos con poco grado. En la Figura 14, si el individuo A quiere transmitir una información a B tardará  $g(A)$  rondas, donde  $g(A)$  es el grado de A, en contactar con B. Lo mismo ocurrirá con B. En cambio el nodo C, al tener un grado de 4, tardará  $g(C)$  rondas en obtener la información de A y  $2 * g(C)$  en pasar esta información a B. Si el grado de A es grande, la información tardará más rondas en llegar a B utilizando su conexión directa que mediante el nodo C, en este caso tardará 12 rondas en pasar la información directamente frente a las 8 rondas que tarda mediante C.

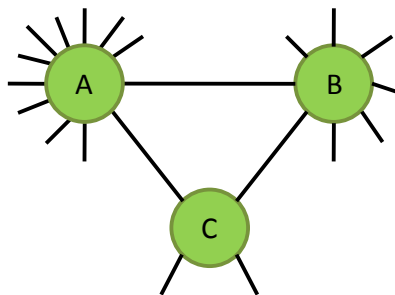


Figura 14 - Ejemplo de red que favorece la propagación de un rumor

Por lo tanto, según observan *Doerr et al.*, uno de los mecanismos que permite que el rumor se propague rápidamente en las redes sociales es que los nodos con menor grado rápidamente adquieren el rumor y lo propagan al siguiente nodo.

#### Propagación y los enlaces débiles

En las conexiones de un grafo podemos distinguir entre dos tipos de enlaces: fuertes y débiles. Los *enlaces fuertes* representan un contacto más cercano y frecuente (amigos), frente a los *enlaces débiles* que muestran una relación más casual (conocidos).

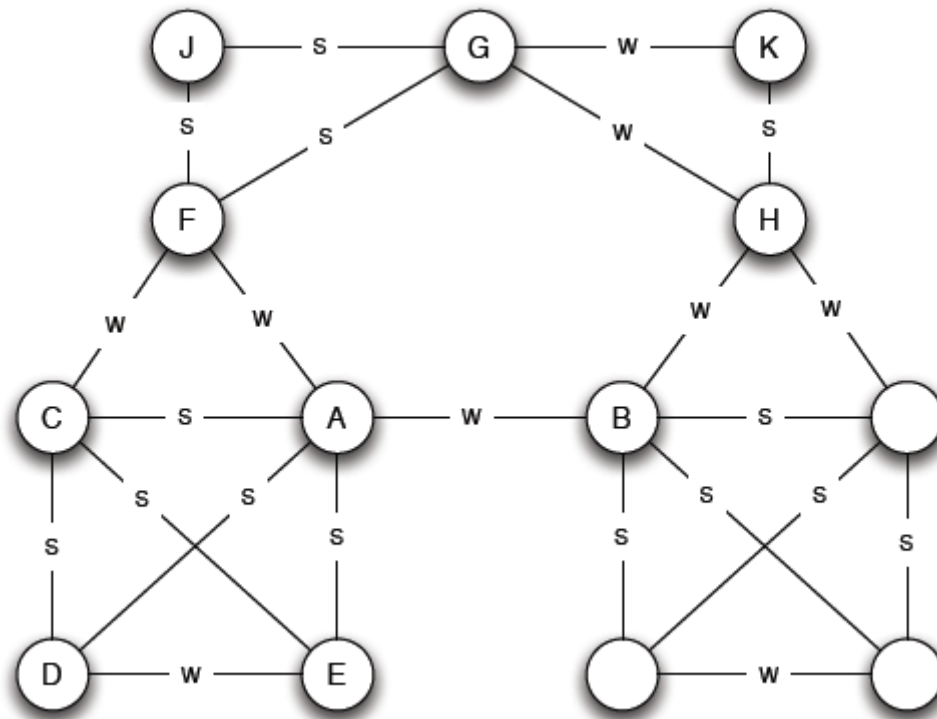


Figura 15 - Enlaces fuertes y débiles. (Kleinberg & Easley, 2010)

En la Figura 15 se muestran los enlaces fuertes (S) y los débiles (W) del grafo. Los enlaces fuertes tienden a estar contenidos en zonas muy conectadas mientras que los débiles suelen conectar diferentes regiones.

Es frecuente que estos enlaces débiles constituyan un medio de acceso a información que, por otro cualquier otro medio, no se pudiese por encontrarse en partes de la red a las que no tiene acceso. Los enlaces débiles constituyen una ventaja puesto que disponen de información que no se encuentra tan relacionada con la información que te proporcionan tus conocidos y que está muy relacionada con la que conoce el individuo.

## 4.2 Influencia en las redes sociales

Cuando las personas se encuentran conectadas en una red, es habitual que influyan el comportamiento y las decisiones de otros individuos. Por ejemplo, en los productos que compran, las actividades que realizan, sus ideas políticas, la tecnología que usan, etc.

Numerosos investigadores han encontrado evidencias de influencia social (Bonchi, 2011): observan, por ejemplo, que los individuos tienden a adoptar comportamientos que han sido realizados previamente por sus vecinos de la red social

Sin embargo, identificar inequívocamente dónde, cuándo, y en qué sentido preciso existe influencia social es más difícil de lo que puede parecer a primera vista, de hecho es todo un reto. No es difícil detectar la existencia de correlación social, es decir, las acciones que realiza un usuario están relacionadas con las acciones de sus vecinos. Sin embargo, esta relación no necesariamente implica una relación de causa efecto entre el vínculo social y la coincidencia en

los comportamientos. En general la correlación puede verse ocasionada por tres motivos principales: homofilia, condiciones externas e influencia social.

### **Homofilia**

La homofilia es el principio por el que los usuarios tendemos a ser similares a nuestros amigos (o dicho de otro modo, a elegir amigos similares a nosotros) y, por lo tanto, a realizar acciones similares. Generalmente la similitud entre personas es un factor causal de la creación de vínculos sociales, más que a la inversa.

Si analizamos nuestra red social de forma agregada, nuestros amigos son frecuentemente similares a nosotros en diferentes aspectos: rango de edad, asistencia al mismo colegio, trabajo, estudiar la misma carrera, vivir en la misma ciudad, mismo barrio, mismos intereses o gustos etc. Estas relaciones son conocidas como relaciones externas.

La homofilia proporciona un ejemplo de cómo el contexto externo condiciona la formación de la red social (Kleinberg & Easley, 2010). Si comparamos el vínculo existente entre dos usuarios con un amigo en común o aquel vínculo consecuente, por ejemplo, de trabajar en el mismo sitio podemos ver el condicionante externo. En el primer caso el enlace viene dado por la estructura de la red: basándonos en el principio de cierre triádico, al que hacíamos referencia en el apartado 2.3, si dos usuarios comparten un amigo la probabilidad de compartir un enlace se incrementa.

Sin embargo, para la emergencia de enlaces por homofilia es necesario prestar atención a las condiciones externas.

### **Condiciones externas**

La realización de una acción por parte de un usuario puede venir ocasionada por causas externas a la red social. Por ejemplo, puede tratarse un evento/noticia que haya llegado de forma externa (por email, visita de páginas) o tratarse de una acción muy popular (adquirir el nuevo iPhone).

### **Influencia social**

Si nos referimos a acciones de los usuarios que puedan ser susceptibles de verse alteradas tales como comportamientos, opiniones, intereses, entre otras, estas pueden verse modificadas para acercarlas a aquellas que sus contactos tienen. Este proceso es el que se denomina socialización o influencia social (Kleinberg & Easley, 2010).

Este caso puede verse como el fenómeno inverso a la homofilia: en el caso de selección, las características individuales de los individuos forman los enlaces. En el caso de influencia social, los enlaces existentes hacen que se modifiquen las características individuales.

Discernir si la actitud de los usuarios tiene su origen en la influencia social es un problema en absoluto trivial. La homofilia y las variables externas pueden establecer una correlación entre las acciones de los usuarios y son por ello numerosos los estudios que tienen su origen en este ámbito. En la práctica cabe suponer que se da una combinación de las tres perspectivas: la similitud entre personas influye en la probabilidad de que se relacione; la comunicación y la

interacción con otras personas modifican las características de los individuos, creando nuevas semejanzas; y obviamente ello se compagina con influencias externas a las relaciones directas entre personas individuales.

Pero, ¿cómo influyen los individuos a sus vecinos para adoptar nuevas ideas o comportamientos? Abordamos esta cuestión en la siguiente sección.

### 4.3 Modelando la difusión en las redes

Para abordar la cuestión anterior es necesario analizar la red y las decisiones individuales que adopta cada uno de los individuos. Analizaremos este proceso desde el punto de vista tanto de la sociología como de los físicos.

#### Sociología

En el apartado 4.2 hemos visto cómo tendemos a relacionarnos con personas similares a nosotros pero sin analizar qué decisiones toman los individuos para ser similares. Para explicar este proceso se ha desarrollado la denominada teoría de la *difusión de innovaciones*, una teoría sociológica que permite explicar cómo, porqué y a qué velocidad se mueven las ideas.

Algunos de los primeros estudios se centraron en los casos en los que la influencia persona a persona procedía de la *información indirecta* que les proporcionaba observar a sus vecinos de la red llevar a cabo esa acción. En este caso, adoptar un producto o realizar una acción tienen su origen en ver cómo sus amigos de la red social la han realizado previamente.

Existen otros modelos basados en *beneficio directo* en lugar de la información indirecta, es decir, tus beneficios en adoptar un nuevo producto se incrementan según tus vecinos lo vayan adquiriendo. En este caso, adoptarás el producto cuando una porción suficiente de tus vecinos lo hayan adquirido previamente. Por ejemplo, si se trata de una red social laboral, será más sencillo trabajar con los compañeros si se trabaja con la misma tecnología.

En base a estos últimos modelos, Kempe *et al.* (2003) proponen los modelos que denomina *Linear Threshold* e *Independent Cascade*, dos modelos de propagación ampliamente estudiados.

En ambos modelos se parte de una red social representada por un grafo dirigido, en la cual cada uno de los nodos individuales puede estar activo (es decir, adopta una innovación) o bien inactivo. Basándose en la motivación anterior, la tendencia de cada nodo a estar activo se incrementa según sus vecinos se convierten en activos. Además, un nodo sólo puede pasar de estar inactivo a estar activo.

En el *Linear Threshold Model* cada nodo  $v$  es influido por sus vecinos  $w$  en función de un peso  $b_{v,w}$  tal que:

$$\sum_{w \text{ vecino de } v} b_{v,w} \leq 1$$

Además, cada uno de los nodos tiene asociado un umbral  $\theta$  comprendido en el intervalo  $[0,1]$  que representa la fracción ponderada de sus vecinos que han de estar activados para que él se active. Este umbral es asignado de forma aleatoria.

El proceso de difusión comienza con un conjunto inicial de nodos activos y transcurre en pasos  $t$  de tal forma que en cada paso  $t$  se activarán los nodos para los que el peso total de la influencia de sus vecinos activos sea al menos  $\theta$ . Es decir:

$$\sum_{w \text{ vecino de } v} b_{v,w} \geq \theta_v$$

Los nodos que en el paso  $t-1$  estuviesen activos continúan estándolo.

El *Independent Cascade Model* comienza, al igual que el modelo anterior, con un conjunto inicial de nodos activos y transcurre en pasos. Cuando un nodo  $v$  pasa a estar activo en un tiempo  $t$ , tienen una oportunidad para activar cada uno de sus vecinos inactivos  $w$ . El éxito viene determinado por una probabilidad  $p_{v,w}$  que es un parámetro del sistema. Si tiene éxito, en el siguiente paso  $t+1$  se activará pero ocurra o no, no podrá intentarlo de nuevo en posteriores rondas.

En ambos modelos, las ejecuciones terminan cuando no se puedan activar más nodos.

La versión determinista del *Linear Threshold* también es conocida como *tipping model* y es asociada con los problemas de maximización de influencia que se explican en la sección Marketing viral.

## Modelos físicos

Durante los últimos años, los físicos estadísticos también han realizado sus contribuciones al estudio de la dinámica en las redes sociales. Uno de los modelos es el propuesto por Lallouache *et al.* (2010) para la formación colectiva de opinión en la sociedad. Este modelo está basado en *Kinetic exchange models*, estudiados en el contexto de economía, y que son modelos dinámicos multi-agente cuyo propósito es explicar las características universales de las distribuciones de la riqueza.

En el modelo propuesto por Lallouache *et al.* asumen que cada persona puede interactuar sólo con una persona en cada instante de tiempo. Este diálogo o discusión es vista como un fenómeno de dispersión en física, en el que las personas se ven como partículas en movimiento que pueden chocar entre sí, al encontrarse, intercambian sus opiniones. La opinión de cada individuo está polarizada, tomando valores comprendidos entre -1 y 1.

$$o_i \in [-1, +1]$$

El intercambio de opiniones entre los individuos  $i$  y  $j$  se realiza como se indica en la siguiente ecuación:

$$o_i(t+1) = \lambda o_i(t) + \epsilon \mu o_j(t)$$

$$o_j(t+1) = \lambda o_j(t) + \epsilon' \mu o_i(t)$$

Es decir, en tiempo  $t+1$  una persona retiene una fracción de su antigua opinión  $o_i(t)$  determinada por el parámetro de convicción  $\lambda$ , característica de cada persona e independiente del tiempo. Al mismo tiempo, cada individuo será influido por la opinión del individuo  $j$  en función del parámetro de influencia  $\mu$ . Las variables  $\epsilon$  y  $\epsilon'$  son números aleatorios uniformemente distribuidos entre 0 y 1 que cambian en cada instante de tiempo.

Lallouache *et al.* asumen que  $\lambda = \mu$ , que sería un caso particular del modelo anterior, y es lo que se conoce como el modelo LCCC:

$$o_i(t+1) = \lambda[o_i(t) + \epsilon o_j(t)]$$

$$o_j(t+1) = \lambda[o_j(t) + \epsilon' o_i(t)]$$

La característica principal de este modelo es que, partiendo de un estado inicial aleatorio en el que las opiniones están uniformemente distribuidas con valores positivos y negativos, converge hasta que existe una opinión media polarizada o consenso.

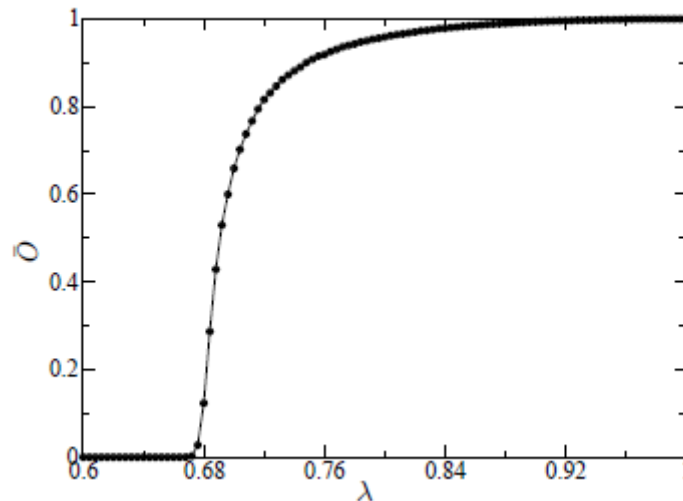


Figura 16 - Variación de la opinión media frente a  $\lambda$  (Modelo LCCC)

La opinión media que se muestra en la Figura 16 ha sido calculada como:

$$\bar{O}(t) = \frac{|\sum_i o_i(t)|}{N}$$

En la Figura 16 que muestra los resultados de las simulaciones realizadas observan que el sistema se estabiliza en dos fases: cuando  $\lambda \leq \lambda_c$  en el que  $\bar{O} = 0$  y cuando  $\lambda > \lambda_c$

$$\begin{cases} \lambda \leq \lambda_c & \bar{O} = 0 \\ \lambda > \lambda_c & \bar{O} > 0 \end{cases}$$

En el caso del modelo LCCC, el valor de  $\lambda_c$  obtenido mediante simulaciones de Monte Carlo es de  $\lambda_c \approx \frac{2}{3}$ . Además, como se aprecia en el caso en el que se supera el umbral crítico ( $\lambda_c$ ), la opinión media crece ( $0 \rightarrow 1$ ) según se toma un valor de  $\lambda$  más cerca de 1.

Este modelo y el efecto de percolación observado es estudiado por otros autores como Chandra (2012). Chandra parte del modelo *square lattice LCCC*, en el que asume que los individuos se encuentran dispuestos en una cuadrícula y siguen la dinámica del modelo LCCC. Sobre este modelo, estudia la transición de percolación de lo que denomina *clusters geométricos*, es decir, sitios adyacentes que tienen un valor de opinión mayor o igual que un umbral prefijado.

El modelo que presentamos en el trabajo guarda relación con la forma en la que los individuos intercambian sus opiniones en el modelo de Lallouache *et al.* Tanto la descripción del modelo propuesto como su relación se encuentran detalladas en el apartado 5.

#### **4.4 Influencia social y sus aplicaciones**

La existencia de influencia procedente únicamente de las redes sociales tiene un gran número de aplicaciones. La más conocida es el marketing viral, en la que etiquetando a los usuarios con mayor posibilidad de influir se puede alcanzar a una proporción muy alta de una red social con muy poco coste.

Entre otras de las posibles aplicaciones se incluyen las recomendaciones personalizadas. Por el principio de homofilia presente en las redes sociales y que hemos estudiado con anterioridad, los usuarios tendemos a relacionarnos con individuos con gustos parecidos a los nuestros. Además, cuando las personas toman decisiones normalmente confían en los consejos que les proporcionan sus amigos o conocidos. Por estos motivos y puesto que los sistemas de recomendación suelen basar su forma de realizar sugerencias sin distinguir si son amigos o no de esta persona, extraer esta información de las redes sociales de los usuarios plantea una posible mejora en la precisión de los recomendadores.

##### **Marketing viral**

Debido a la influencia social y cómo puede verse modificado el comportamiento o la opinión de los usuarios, surge el concepto de marketing viral. Este concepto es empleado para referirse a las técnicas que intentan explotar redes sociales y otros medios electrónicos para difundir su mensaje. Se suelen basar en el proceso de difusión mediante el “boca a boca”.

Uno de los problemas clave en estas técnicas es identificar los usuarios claves a los que dirigir el mensaje para maximizar la difusión. El problema es detectar qué nodos iniciales han de ser elegidos para la propagación con el fin de *maximizar la influencia*.

Los primeros en abordar el problema y de detectar aquellos usuarios a los que difundir el mensaje fueron Domingos y Richardson (2001). Posteriormente fue abordado por Kempe *et al.* (2003). Si se define la influencia esperada del conjunto  $S$ , denotada por  $\sigma_m(S)$ , como el conjunto de nodos activos al final del proceso siendo  $S$  el conjunto inicial de nodos, el problema de maximización de influencia consiste en seleccionar el conjunto  $S$  tal que la influencia esperada sea máxima.

El problema de maximización de influencia abordado tanto desde el *Threshold Model* como del *Independent Cascade Model*, explicados en la sección 4.3, es NP-Hard. Kempe mostró que la función  $\sigma_m(S)$  es monótona y submodular lo que permitía obtener una aproximación gracias a un algoritmo avaro. Monótona implica que según crece el conjunto de nodos activados, la probabilidad de que un nodo se active no debe disminuir. Submodular se refiere a que la ganancia obtenida al añadir un nodo al conjunto  $S$  es mayor o igual que añadirla al súper conjunto de  $S$ . Es decir, la probabilidad de activar un nodo inactivo no se incrementa si más nodos han tratado de activarlo previamente. El conjunto obtenido con este algoritmo garantiza tener una difusión  $> 63\%$  de la difusión esperada por el conjunto óptimo.

Puesto que no se puede aplicar a redes reales debido a que es necesario ejecutar muchas simulaciones de propagación para obtener el resultado, numerosos investigadores se han centrado en mejorar la eficiencia y la escalabilidad de los algoritmos de maximización de influencia. Aproximaciones como la de (Shakarian & Paulo, 2012) garantizan encontrar un conjunto de nodos, no necesariamente mínimo, que si se seleccionan como iniciales el mensaje llegará a toda la red y, además, con buena escalabilidad. Este algoritmo se basa en la idea de shell-decomposition en el que los vértices son eliminados del grafo, en función del grado que tengan, hasta que se obtiene el conjunto resultante.

En el marketing viral se asume que el grafo de influencias es proporcionado como entrada del problema. Sin embargo, las redes sociales reales no vienen con aristas etiquetadas con las probabilidades de influencia. En la realidad vienen con un log de acciones pasadas, es decir, una tabla en la que se muestran de forma cronológica las acciones llevadas a cabo por los usuarios.

En base a esta idea, Goyal *et al.* (2010) desarrollan una serie de algoritmos que, tomando como base el grafo y el log de acciones, permiten aprender los parámetros de los modelos con los que obtener las influencias individuales y para probar los modelos.

Los modelos propuestos por Goyal para estimar la probabilidad de influencia se clasifican en 3 tipos:

Modelos estáticos → aquellos que no dependen del tiempo. Como ejemplos están las distribuciones de Bernoulli o el índice de Jaccard.

Modelos continuos de tiempo → Son aquellos en los que se basan en que la probabilidad de influir no permanece constante en el tiempo. Asumen que, cuando un usuario observa a un amigo realizar una acción, tiene más necesidad de propagarla. Esta influencia va decayendo con el tiempo.

Modelos discretos de tiempo → Proponen una aproximación de los modelos continuos. En éstos, la probabilidad de influir permanece constante durante una ventana de tiempo, tras la cual cae a 0.

En el análisis realizado los autores observan que las pruebas sobre los modelos discretos son más eficientes que los modelos continuos (cuyo coste para probarlos es muy elevado) y con ellos se pueden obtener niveles de precisión cercanos a los continuos en el tiempo.



## **Recomendación personalizada**

El propósito de los sistemas de recomendación es ofrecer a los usuarios recomendaciones que puedan ser de su interés. Estos sistemas pueden beneficiar tanto a los clientes como a las empresas: pueden reducir el tiempo y el coste de buscar ítems online, proporcionando en ocasiones ítems no esperados. También facilitan el uso de las aplicaciones comerciales, hecho que puede retener a los clientes e incrementar las ventas. Algunos ejemplos de estas aplicaciones incluyen la recomendación de libros, CDs y otros productos en Amazon.com, películas en Movilens y noticias en VERSIFI Technologies.

Los sistemas de recomendación toman como entrada preferencias de los usuarios e inversamente les influyen pues adquieren una intervención (mayor o menor) sobre sus decisiones y sus experiencias, que son lo que conforma los gustos de las personas. Así pues es de interés preguntarse cómo influyen las derivas de opinión, condicionadas por procesos de comunicación y “contagio” de gustos y experiencias entre personas, en el comportamiento de un sistema de recomendación que opera en ese entorno. Y a la inversa, un sistema de recomendación se puede convertir en un agente más que influye y provoca la propagación de cambios de preferencia.

Los sistemas de recomendación usan diferentes algoritmos para filtrar los datos y generar las recomendaciones a los usuarios. Adomavicius (2005) clasifica en 3 los tipos de recomendadores, atendiendo a la metodología que siguen al realizar las recomendaciones:

- *Recomendación basada en contenido:* Se recomendarán ítems al usuario que sean similares a aquellos que haya preferido en el pasado.
- *Filtrado colaborativo:* En este caso se recomendarán ítems en base a la experiencia de otros usuarios con gustos o rasgos similares.
- *Recomendación híbrida:* Estos métodos combinan los dos métodos anteriores.

Como hemos visto en secciones anteriores, la red social de una persona es determinante en sus gustos y preferencias. En el apartado 4.2 hemos visto cómo los usuarios tendemos a ser similares a nuestros amigos y, por lo tanto, a realizar acciones similares. Por este motivo, disponer de información de la red de las personas puede ayudarnos a inferir sus características o gustos.

Estas similitudes pueden estar causadas por influencia y las interacciones de la red social: las personas tienden a recordar la información que le proporcionan sus amigos más que la información abstracta. Por lo tanto, la toma de decisiones por parte de los usuarios se encuentra condicionada a las opiniones de sus amigos y personas más cercanas ya que tienden a confiar en ellos. En un experimento realizado por (Sinha & Swearingen, 2001) en el que comparaban la calidad de las recomendaciones ofrecidas por los sistemas como Amazon y recomendaciones de sus amigos encontraron que estas últimas eran las preferidas por los usuarios.

Una desventaja de métodos como el filtrado colaborativo es que no son capaces de distinguir a los vecinos como amigos o extraños con gustos similares. Por este motivo y gracias a las redes sociales online que permiten facilitar la tarea de identificar a los amigos de los usuarios,

autores como Liu y Lee (2010) han realizado aproximaciones combinando ambas técnicas. Proponen un método que combina los vecinos próximos del filtrado colaborativo (método para calcular el nivel de similitud de los usuarios, y la información procedente de las redes sociales. En su estudio observan que este tipo de métodos son más efectivos, mejorando la precisión de su predicción.

En este trabajo, como presentamos en la sección 6.6, estudiamos la introducción de un agente recomendador en el sistema para ver su efecto en la dinámica de formación de opiniones.

## 5. Descripción del modelo

### 5.1 Estructura social

En el proceso de formación de opinión partimos de un conjunto de individuos (representados como nodos del grafo), lo cuales se relacionan con otros individuos de la misma red. Esta relación se muestra con una arista. Consideramos redes no dirigidas, esto es, las aristas reflejan relaciones bidireccionales. Un ejemplo de este tipo de red son las relaciones de amistad en Facebook, donde dos usuarios, si son amigos, acceden a aquellas ciertas publicaciones restringidas a este tipo de visibilidad. En el estudio asumimos que la velocidad a la que la estructura de la red cambia es mucho más lenta que el proceso de formación de opinión por lo que, para todo el experimento, consideramos invariable la estructura del grafo a lo largo del tiempo.

### 5.2 El modelo de opinión

Cada usuario dispone de un conjunto de ítems de los cuales tiene una opinión formada. Por ejemplo, pueden representar opiniones de películas, de libros, sobre temas de actualidad, o artículos de consumo.

Esta opinión se encuentra distribuida con valores comprendidos entre 0 y 1, donde valores más altos implicarán una mayor preferencia por el objeto. La asignación de opiniones se realiza al iniciar el sistema de forma que, al comenzar la simulación, cada usuario dispondrá de un conjunto de ítems y de su opinión al respecto. Los valores se pueden modificar posteriormente por efecto de la comunicación entre usuarios, como describiremos más abajo, modelando el correspondiente proceso de influencia que tiene lugar cuando dos personas se comunican sus opiniones.

Con el fin de no disponer de valores independientes, las opiniones son normalizadas cada vez que se modifican, aplicando la fórmula:

$$O_u^{k'}(t) = \frac{O_u^k(t)}{\sum_{i=0}^I O_u^i(t)}$$

Donde

- $O_u^{k'}(t)$  representa la opinión normalizada del usuario  $u$  sobre el ítem  $k$  en un tiempo  $t$ .
- $O_u^k(t)$  representa la opinión del usuario  $u$  sobre el ítem  $k$  en un tiempo  $t$ .

Así, la Figura 17 mostraría un ejemplo de opiniones de 5 ítems para un usuario, y su consiguiente normalización.

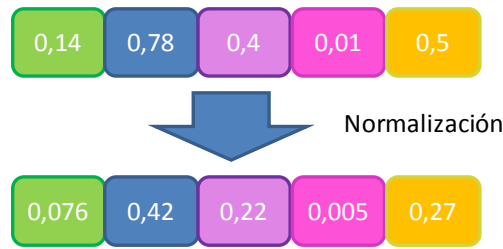


Figura 17 - Normalización de las opiniones

En este caso, el usuario prefiere el ítem azul frente al resto. Esta normalización se repite cada vez que se calculan las nuevas opiniones, de forma que todas las opiniones que cada usuario tiene de los ítems siempre suman uno. La normalización tiene el efecto colateral intencionado de crear una dependencia entre las opiniones que tiene cada usuario de los ítems, puesto que si mejora la opinión de un usuario sobre un ítem, descende la preferencia por los demás al normalizar. Esta dependencia permite desarrollar un modelo más interesante del que se obtendría en caso contrario, pues de no haber tal dependencia estaríamos trabajando con  $I$  simulaciones paralelas e independientes que en definitiva se reducen a réplicas de un experimento individual, donde  $I$  representa el número de ítems de los que se tiene opinión.

### 5.3 El modelo de interacción

Para la formación de la opinión de los individuos definimos un modelo basado en agentes que consiste en una modificación del modelo de Lallouache *et al.* (2010) para simular las dinámicas colectivas de opinión. Las modificaciones aplicadas permiten adaptar el escenario a múltiples ítems, pasando a una única opinión al estudio de la dinámica de  $I$  opiniones.

El comportamiento individual de un usuario viene determinado por 3 parámetros:

- La convicción que tengo en mi opinión, representada como  $\lambda$  ( $0 \leq \lambda \leq 1$ ). Este parámetro muestra cuánto de convencido está el usuario de su opinión.
- Cuánto me dejo influir, representado como  $\mu$  ( $0 \leq \mu \leq 1$ ). Muestra en qué medida cada persona se deja influir por la opinión de su vecino. Para el experimento se asume que la influencia es  $(1 - \lambda)$ . Valores próximos a 0 implican que los individuos no se dejan influir con facilidad y están muy convencidos de su opinión mientras que valores próximos a 1 implican que la persona es altamente influenciable por lo que vea en las opiniones de sus amigos.
- La influencia exógena, representada como un parámetro  $\alpha \in [0,1]$ . Muestra opiniones adicionales que le llegan al individuo de forma externa a las personas de su entorno como, por ejemplo, por vía de los medios de comunicación. Este parámetro se podría representar también como los defectos (de expresión y entendimiento) en la comunicación, si bien a los efectos del presente estudio resulta de más interés el sentido del factor externo.

Estos parámetros influyen en la nueva opinión del individuo en el instante  $t+1$  como muestra la siguiente fórmula:

$$O_u^i(t+1) = \lambda_u O_u^i(t) + \mu_v O_v^i(t) + f(\alpha)$$

Donde,

- $O_u^i(t + 1)$  representa la opinión sobre el ítem  $i$  por parte del usuario  $u$  en el instante  $t + 1$ .
- $O_u^i(t)$  representa la opinión sobre el ítem  $i$  del usuario  $u$  en el instante  $t + 1$ .
- $\lambda_u$  es la convicción del usuario  $u$  en su propia opinión.
- $O_v^i(t)$  representa la opinión sobre el ítem  $i$  del usuario  $v$  en el instante  $t + 1$ , donde  $v$  es vecino de  $u$ .
- $\mu_v$  representa la capacidad de persuasión de  $v$ .
- $\alpha$  representa la influencia exógena.

A fin de evitar que las opiniones se salgan de su rango por efecto de la influencia externa, el valor de  $\alpha$  es matizado aplicando la siguiente fórmula:

$$f(\alpha) = O_u^i(t + 1)' * \alpha * ((2 * rand(0,1) - 1))$$

Donde,

- $O_u^i(t + 1)'$  representa el valor de la nueva opinión del usuario sin tener en cuenta la influencia exógena ( $\lambda_u O_u^i(t) + \mu_v O_v^i(t)$ ).
- $\alpha$  representa el porcentaje del factor exógeno que aplicar con respecto al nuevo valor de opinión. Este parámetro será el que modificaremos para ampliar o reducir la influencia exógena que introduzcamos en el sistema.

Es decir, en cada interacción el usuario retiene parte de su antigua opinión  $O_u^i(t)$  determinada por el valor de su convicción  $\lambda_u$  y es influido por la opinión de la otra persona  $O_v^i(t)$  en función de su persuasividad  $\mu_v$ . Para simplificar, consideramos que tanto la influencia como la convicción son iguales para todos los nodos de la red, asumiendo que no existen líderes en el grafo. Bajo esta simplificación podemos, además sin pérdida de generalidad, reducir  $\lambda$  y  $\mu$  a un solo parámetro con  $\mu = 1 - \lambda$ .

## 5.4 Dinámica de propagación

En cada paso de la simulación cada usuario va a interactuar con sus vecinos, modificando su opinión en base a la información que le llega de las opiniones del resto de individuos. Es decir, el usuario  $u$  en cada instante  $t$  va a conocer la opinión de uno de sus vecinos acerca de un ítem, influyendo en la opinión que tiene al respecto. En cada interacción, únicamente se modifica la opinión del usuario  $u$  que es influido, escenario similar a leer la opinión de un amigo de un amigo sobre un objeto en Internet.

Por ejemplo, el grafo puede representar una red social online donde los enlaces muestran los amigos de los usuarios. Los ítems podrían ser los cinco últimos estrenos de películas de los cuales la gente se encuentre opinando. El usuario, haya visto o no las películas, ya dispone de una opinión acerca de las mismas pero, al leer lo que opinan sus amigos de ellas, influye en su propia opinión. Esto ocurre para todos los usuarios del grafo, ya que se encuentran leyendo las opiniones del resto de sus amigos. El propósito es ver cómo va evolucionando la opinión del conjunto mediante las influencias individuales que ejercen las opiniones del resto de la red.

El modelo permite también reflejar otros escenarios donde dos personas se encuentran (online u offline) y se comunican bidireccionalmente. Ello se reflejaría en nuestro modelo estableciendo una influencia doble en la interacción entre dos vecinos. En lo sucesivo nos

centramos en la opción unidireccional, si bien hemos podido comprobar experimentalmente que la opción bidireccional produce resultados equivalentes.

La dinámica seguida en cada uno de los pasos de la simulación para modelar el comportamiento de cada individuo se define en el pseudocódigo de la Figura 18.

```
for all Users in Graph

    usuariosVecinos <- dameVecinosDelUsuario(user);
    if usuariosVecinos.length > 0

        for all Items in Items

            opinionNueva(user)[item] = calculaNuevaOpinionDelItem();

        end for
        opinionNueva(user) = normalizaOpiniones(opinionNueva);

    else
        opinionNueva(user) = opinionActual(user)
    end if

end for

// AL FINAL DE LA SIM. SE PERSISTEN
// LAS NUEVAS OPINIONES
for all Users in Graph

    opinionActual(user) = opinionNueva(user)

end for
```

Figura 18 - Pseudocódigo de la dinámica de propagación

En cada tiempo  $t$  se recorren todos los nodos del grafo. Para cada usuario, se consulta si tiene o no vecinos: en caso de no tener, la opinión en  $t+1$  del individuo coincidirá con que tiene en el tiempo  $t$ . Si el usuario  $u$  tiene vecinos, se calcula, para cada uno de los ítems la nueva opinión que tendrá en el tiempo  $t+1$  en base al modelo descrito en el apartado 5.2.

Una vez finalizados los cálculos de las nuevas opiniones en tiempo  $t+1$  de todos los usuarios, estas opiniones pasan a ser las actuales, avanzando un paso en la simulación. De esta forma, evitamos que el orden en el que se seleccionen los individuos del grafo influya en la dinámica de formación de opinión.

A continuación se muestra en la Figura 19 el proceso que tiene lugar en cada tiempo  $t$  para cada uno de los individuos del grafo.

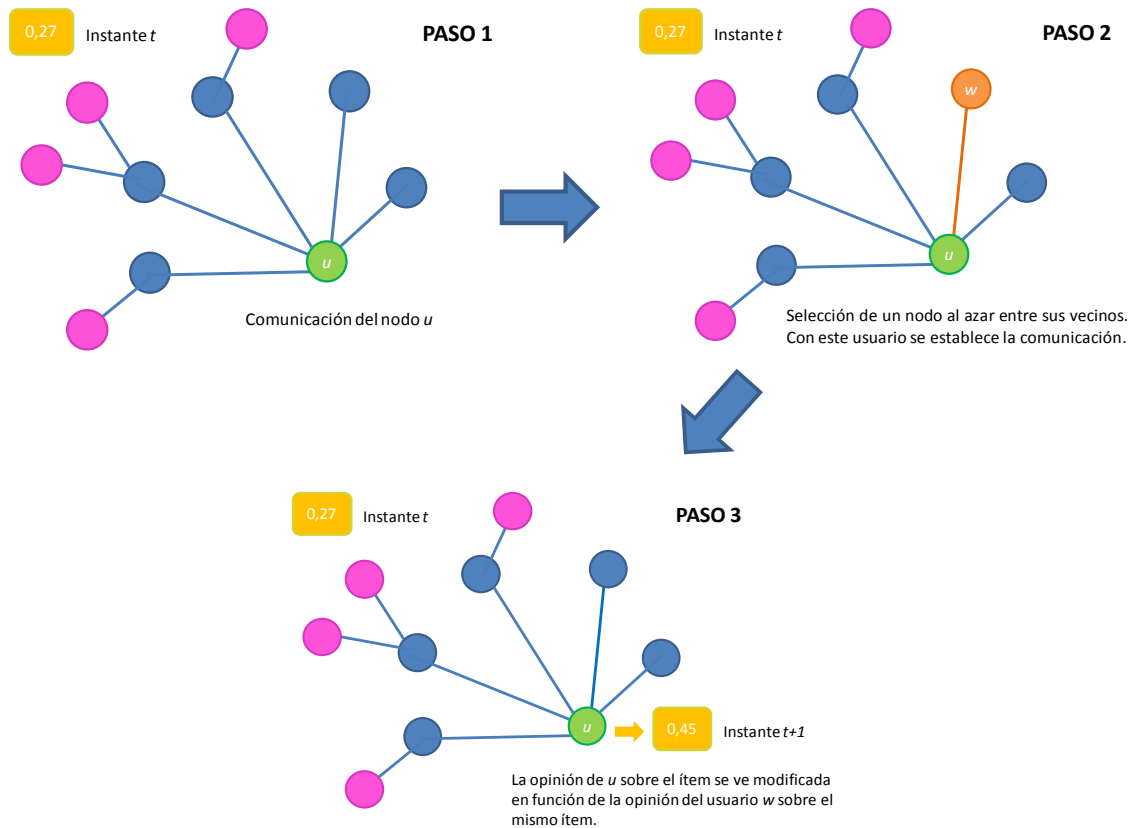


Figura 19 - Secuencia del algoritmo

1. En el paso 1 se muestra el individuo (en color verde) que va a informarse sobre uno de los ítems, en este caso el de color amarillo. El número contenido dentro del ítem indica la preferencia del individuo  $u$  por el ítem. Este proceso se repetiría para todos sus ítems.
2. A continuación se selecciona, de forma aleatoria, la persona cuya opinión sobre el ítem el usuario activo va a conocer. Este usuario tiene que ser un vecino de  $u$  (en naranja en la Figura 19). Un ejemplo de este proceso podría ser que el individuo  $u$  viese una entrada en el muro del individuo  $w$  opinando acerca del ítem.
3. Por último, la opinión de  $u$  se ve influida por la opinión que tenga  $w$  del ítem, modificándose positiva o negativamente, de acuerdo con las ecuaciones del modelo definidas en la sección anterior. Esta nueva opinión tendrá efecto en el instante  $t+1$ .

Este proceso es repetido en cada uno de los pasos para cada usuario y para todos los ítems ( $i \in [1 \dots I]$ ), de tal forma que la opinión que tiene cada uno de los individuos de cada ítem es modificada en cada paso de la simulación. La actualización con los nuevos valores de opinión así calculados se lleva a efecto una vez que se hayan calculado las nuevas opiniones de todos los ítems, es decir, al final de cada paso. De esta forma, el orden en el cual se recorran los nodos no influye en el resultado.

## 5.5 Metodología de la recomendación

En el modelo, vamos a introducir un agente externo al grafo que realice recomendaciones a los usuarios para estudiar el efecto que tiene en la formación de la opinión.

Para introducir un nuevo agente recomendador, es necesario modificar la forma en la que los usuarios interactúan con sus vecinos. Se define la probabilidad  $p^r$  como la probabilidad de recibir información de un recomendador. Por lo tanto, en nuestro sistema, con probabilidad  $p^r$  el individuo se verá influido por lo que le recomienden, mientras que con probabilidad  $1 - p^r$  se verá influido por la información de los ítems que le proporcionen sus vecinos.

En nuestro caso, debido a las características del escenario de la simulación, trabajaremos con un recomendador basado en popularidad. Es decir, en cada momento  $t$  evaluará qué ítem es el más popular entre los usuarios del grafo, asignando un 1 a su opinión.

El proceso de formación de opinión se realiza de la misma forma que si la interacción se realizase con otro usuario:

$$O_u^i(t+1) = \lambda_u O_u^i(t) + \mu_v O_v^i(t) + f(\alpha)$$

Donde,

- $O_v^i(t)$  representa la opinión del recomendador sobre el ítem  $i$  en el instante  $t$ .

En este caso, si el usuario interactúa con el recomendador, únicamente modificaría la opinión del ítem más popular y no modificaría el resto de opiniones.

## 5.6 Trabajo relacionado

El modelo que proponemos y exploramos en este trabajo está relacionado con los modelos de propagación de epidemias analizados en el área de las dinámicas de red 3.2. Una primera diferencia entre estos modelos y el nuestro es la binaridad del estado del usuario en aquéllos frente a la pluralidad en el nuestro. Se puede establecer por tanto una con nuestro modelo donde la propagación de epidemias sería un caso particular con dos ítems (infectado y sano). Por otra parte la función de actualización del estado en el contacto entre dos individuos es diferente en ambos casos: típicamente la “influencia” es binaria bajo una determinada probabilidad Bernoulli en el caso de los modelos de epidemias, mientras que en nuestro caso se trata de una función escalar sobre los grados de preferencia de cada ítem, en base a cuyo resulta la elección de una opción preferida (un estado –un ítem– resulta seleccionado).

Siguiendo esta línea de conexión, podrían estudiarse variaciones de nuestro modelo tales como introducir el concepto de inmunidad (modelos *SIR*) en forma de individuos sobre los que el resto de personas no ejerce influencia, similar al rol del líder, o dar la posibilidad de perder esta inmunidad (modelos *SIRS*) y permitir que comiencen a ser influidos.

Por lo demás, el estudio de cómo se comportan los individuos de forma colectiva es un campo donde numerosos investigadores han realizado sus aportaciones. Wu y Huberman (2004) proponen una teoría de formación de opinión en la que incluyen elecciones de los individuos entre 2 ó 3 opiniones, prediciendo su evolución en una red social. Aunque también modelan estas opiniones como estados, éstos representan valores discretos: hablan de opiniones “binarias” blanco o negro como los posibles colores que pueden tener las opiniones. Sería equivalente a tener una opinión igual a, por ejemplo, 1 (color blanco), -1 (color negro) y 0 (color



azul). No consideran pues opiniones sobre ítems distintos, sino diferentes valores de opinión sobre el mismo tema. En este sentido, nuestro modelo es una generalización de aquél, donde en lugar de un valor discreto las opiniones son escalares en un intervalo continuo  $[0,1]$ , pero además se maneja un vector de estos valores, y las variaciones en un valor altera los demás para el mismo usuario (por efecto de mantener la suma en 1 para cada individuo), o dicho de otro modo, se consideran opiniones relativas (mejores o peores en referencia a la opinión sobre las demás opciones).

Hammer *et al.* (2012) aplican modelos de dinámica de opinión para investigar la formación de opinión en comunidades en función de la estructura de la red. Al igual que en nuestro modelo, trabajan con valores de opinión comprendidos entre 0 y 1. Además, incluyen un valor de tolerancia, que representa cuánto de receptivo se encuentra el individuo a las opiniones de los demás. Este parámetro regula que las personas sólo se dejen influir por gente con una opinión que no diste de la suya más que el valor de tolerancia fijado y permite la emergencia opiniones por clústers. La opinión del individuo es modificada añadiendo la media de opiniones de sus vecinos, es decir, en cada paso el individuo es influido por la opinión que tengan todos sus vecinos más similares a él en opinión. De nuevo en el estudio se considera la dinámica de una única dimensión de preferencia.

Al igual que en el modelo de Hammer *et al.*, en nuestra propuesta en cada instante de tiempo  $t$  un individuo puede recibir influencia de varios vecinos pero, en nuestro caso, sobre ítems distintos. Suponiendo una red social donde cada usuario comenta, por ejemplo, los últimos estrenos de películas, nuestro modelo asume que, sobre cada película, los individuos únicamente comparten una opinión en cada instante de tiempo  $t$ , siendo influidos positiva o negativamente. La influencia relativa a un ítem repercute no obstante en la preferencia por los demás ítems, pues para mantener constante la suma, si la opinión sobre un ítem mejora para un usuario, decrece su peso de preferencia por los demás.

Nuestro modelo se basa de las funciones de interacción de Lallouache *et al.* (2010) para simular las dinámicas colectivas de opinión, modelo en el que también se basan autores como Chandra (2012). El intercambio de opiniones entre los individuos  $i$  y  $j$  se realiza en base a la siguiente ecuación,

$$o_i(t+1) = \lambda o_i(t) + \epsilon \mu o_j(t)$$

Con valores de  $o_i \in [-1, 1]$ . Nuestro modelo comparte con el de Lallouache *et al.* aspectos como el concepto de convicción  $\lambda$  y persuasión  $\mu$ ,

$$O_u^i(t+1) = \lambda_u O_u^i(t) + \mu_v O_v^i(t) + f(\alpha)$$

Sin embargo, no introducen la interdependencia entre ítems considerando, de nuevo, una única dimensión de preferencia.

## 6. Estudio experimental

### 6.1 Propósito

El propósito del estudio realizado es analizar el proceso de formación de opinión en las redes a través de dinámicas de influencia y propagación entre usuarios. Para ello, se ha realizado un programa que permite analizar el efecto sobre diferentes tipos de grafos y con diferentes configuraciones de parámetros, tanto relativos al comportamiento de los usuarios en su intercomunicación, como a las condiciones iniciales de la que parte el proceso interactivo. El programa permite tanto la ejecución de dicho proceso como la visualización del grafo y de la evolución de la opinión de cada individuo paso a paso, permitiendo un análisis cualitativo de los casos. Junto con esto, la herramienta recopila datos agregados a través del tiempo de simulación para su estudio cuantitativo.

### 6.2 Descripción de las pruebas

El propósito principal de las pruebas es analizar los efectos que tienen los parámetros de convicción, densidad del grafo y la influencia exógena en la formación de opinión en la red. Variando estos parámetros se observará el efecto que tiene en variables como el porcentaje de simulaciones que convergen a una opinión unánime, el tiempo que tarda en llegar a este estado o el número de usuarios que prefiere cada ítem.

Los casos de estudio se han realizado sobre un grafo real y dos tipos de grafo aleatorios: Erdős-Rényi y Barabási-Albert. El grafo real consiste en la unión de la red ego de 10 usuarios de Facebook, disponible en la colección SNAP de Stanford<sup>1</sup> y utilizada en (McAuley & Leskovec, 2012). Para cada medición se ejecutan 20 simulaciones y se promedian las observaciones.

Exceptuando las pruebas realizadas para ver el efecto al variar la densidad (realizadas con 1000 usuarios), las características de los grafos utilizados en el resto de pruebas se muestran en la siguiente tabla.

TIPO DE GRAFO	Parámetros	# Vértices	# Aristas	# Ítems (I)
Erdős-Rényi	Probabilidad = 0.01	4000	---	10
Barabási-Albert	Grado = 40	4000	79600	10
Real	---	3959	84243	10

Tabla 2 - Características de los grafos utilizados en las pruebas

Como se ve en la tabla, los grafos tienen características similares en cuanto a número de vértices y densidad. Para todas las pruebas, se realiza un análisis de los resultados obtenidos atendiendo a las variables y parámetros que se describen en los siguientes apartados.

### 6.3 Variables estudiadas

En esta sección se describen las variables dependientes que vamos a analizar en nuestros estudios, indicando, además, cómo se calculan. En la Tabla 3 se muestra el resumen de estas variables junto con los parámetros que modificaremos para observar su comportamiento.

<sup>1</sup> Fuente: <http://snap.stanford.edu/data/>

Variables dependientes	Parámetro
Usuarios por ítem preferido	Convicción ( $\lambda$ ), densidad e Influencia exógena ( $\alpha$ )
Porcentaje de convergencia	
Porcentaje de usuarios en ítem más popular	
Tiempo de convergencia	
Número medio de ítems preferidos	

Tabla 3- Variables y parámetros de estudio

### Usuarios por ítem preferido

- **Nombre:** Preferencia de ítem.
- **Descripción:** Representa el número de usuarios que prefieren cada ítem en cada paso de la simulación.
- **Cálculo:** La preferencia de un ítem  $i$  en un instante de tiempo  $t$  se calcula como

$$P^i(t) = \sum_{u=1}^N p_u^i(t)$$

Donde la preferencia de cada usuario viene dada por la siguiente fórmula.

$$p_u^i(t) = \begin{cases} 1 & \text{Si } O_u^i(t) > O_u^j(t), \forall j \in I, j \neq i \\ 0 & \text{resto} \end{cases}$$

Es decir, en cada uno de los pasos de la simulación se recorren todos los usuarios del grafo contabilizando qué ítem es el preferido de cada uno de ellos (aquel cuya probabilidad de ser elegido sea mayor).

### Porcentaje de convergencia

- **Nombre:** % de convergencia.
- **Descripción:** Este parámetro hace referencia a la cantidad de las pruebas realizadas que han convergido a una opinión uniforme de todos los usuarios del grafo.
- **Cálculo:**

$$C = \frac{\sum_{s=1}^S c(s)}{S}$$

Donde  $S$  indica el número de simulaciones (en nuestro caso 20) y  $c(s)$  indica si la simulación converge.

$$c(s) = \begin{cases} 1 & \text{Si } \exists i \in I \mid \frac{P^i(t = \text{end})}{N} > 0.99 \\ 0 & \text{resto} \end{cases}$$

Es decir, se determina que una simulación converge si en el último paso de la simulación el 99% de los usuarios prefieren el mismo ítem.

### Porcentaje de usuarios en ítem más popular

- **Nombre:** % usuarios en ítem más popular.
- **Descripción:** Indica el porcentaje de usuarios que prefiere el ítem más popular cuando no converge la simulación. Proporciona una idea de qué porcentaje del grafo tiene la misma opinión.
- **Cálculo:** Para obtener el valor, se mira cuál es el ítem cuya preferencia es mayor al final de la simulación ( $P^i(t = end)$ ), viendo cuánto representa del total de usuarios.

$$\%U^i = \frac{P^i(t = end)}{N} \mid P^i(t = end) > P^j(t = end), \forall j \in I, j \neq i$$

Donde  $N$  indica el número de usuarios del grafo.

### Tiempo de convergencia

- **Nombre:** Tiempo de convergencia.
- **Descripción:** Indica en qué paso ha convergido la simulación, en caso de hacerlo. Permite estimar la velocidad a la que se forma la opinión del grupo.
- **Cálculo:**

$$t(s) = \begin{cases} numStep & \text{si 's' converge a un ítem} \\ 0 & \text{en caso contrario} \end{cases}$$

Para ver el número del paso en el que tiene lugar la convergencia se busca cuál es el primer paso en el que ocurre que  $c(s) = 1$ .

### Número medio de ítems preferidos

- **Nombre:** #medio ítems preferidos
- **Descripción:** Indica cuántos ítems son preferidos, en media, cuando finalizan las simulaciones. Permite ver cómo de homogéneas / heterogéneas son las opiniones de los usuarios del grafo: si únicamente prefieren 2 ó 3 ítems, o si todos los ítems son los preferidos por un porcentaje del grafo.
- **Cálculo:** Para calcularlo, en base al número de usuarios que prefieren cada ítem al final de la simulación, nos quedamos con los que tienen mayor valor (cuya suma de usuarios que lo prefieran lleguen al 95% de los usuarios). De esta forma obtenemos en media, cuántos ítems se quedan compitiendo.

Determinar el 95% permite eliminar de la lista aquellos ítems cuya preferencia es minoritaria.

## 6.4 Arquitectura del programa

El programa se ha implementado en Java utilizando MASON (Multi-Agent Simulator of Networks<sup>2</sup>), conjunto de herramientas de simulación multiagente, y JUNG (Java Universal Network/Graph Framework<sup>3</sup>) librería para el análisis, modelado y visualización de datos que pueden ser representados como un grafo o una red. La Figura 20 muestra el esquema de la aplicación.

<sup>2</sup> <http://cs.gmu.edu/~eclab/projects/mason/>

<sup>3</sup> <http://jung.sourceforge.net/>

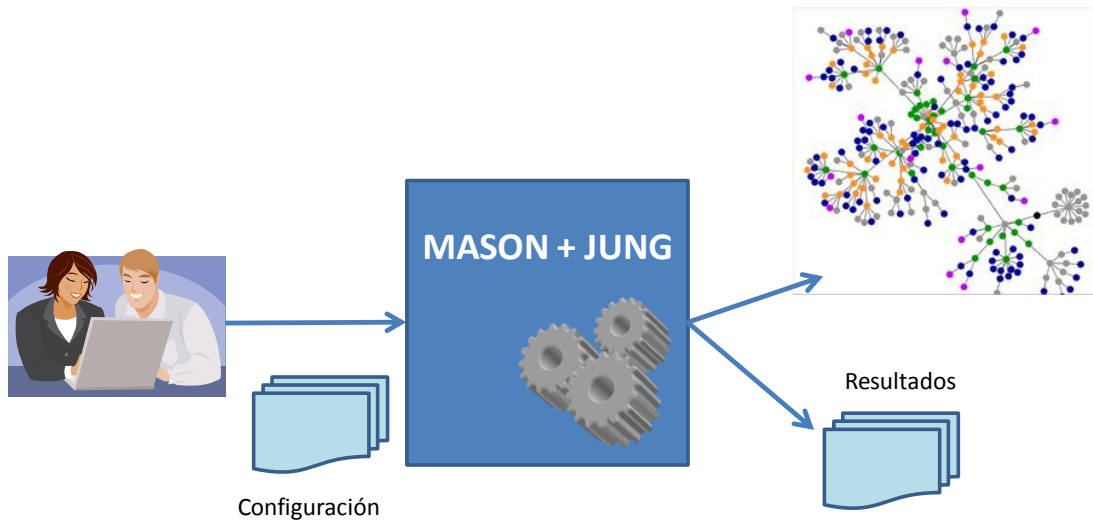


Figura 20 - Esquema de la aplicación

El sistema permite trabajar tanto con datos de grafos reales o de cualquier otro tipo, suministrados en un archivo. Adicionalmente permite generar grafos aleatorios en el propio programa a partir de los parámetros que se definan en un archivo de configuración. Además, permite visualizar el grafo y el proceso de formación de opinión según se produce.

La Figura 21 muestra un ejemplo de cómo se visualiza en el sistema un grafo de 100 nodos.

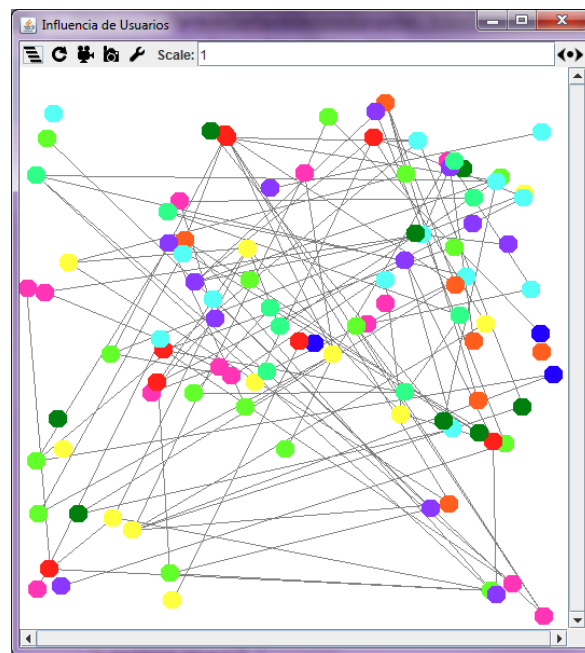


Figura 21 - Ejemplo de grafo visualizado con Mason

En la ilustración se pueden ver nodos que no tienen enlaces y, por lo tanto, no se relacionan con ningún individuo del grafo. Además, cada uno de los nodos está representado por un color distinto. El color muestra qué ítem es el que prefiere el usuario en cada momento de la simulación. Por ejemplo, si la Figura 22 representa el conjunto de opiniones de un usuario sobre los ítems

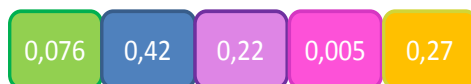


Figura 22 - Opiniones de un usuario

En la visualización del estado del nodo que representa al usuario estaría pintado de color azul, implicando que en ese momento es el ítem que prefiere.

## 6.5 Análisis cualitativo

Describimos en primer lugar a nivel general la casuística observada en las derivas del proceso de formación de opiniones, basada en una primera constatación de casos y tendencias particulares, así como la observación de ejecuciones individuales y ejemplos reducidos. Posteriormente presentaremos los resultados de un estudio más sistemático de las dependencias observadas entre variables del sistema.

### Preferencia por el mismo ítem

Al referirnos a preferencia por el mismo ítem nos encontramos en un estado en el que, partiendo de un estado inicial en el que las opiniones son aleatorias y, por lo tanto, cada usuario preferirá un ítem distinto, todos los individuos de la red acaban prefiriendo el mismo ítem tras un número variable de iteraciones. Esta preferencia es estable y continua en el tiempo.

En la Figura 23 se muestra la evolución de las preferencias de los usuarios en un grafo Erdős-Rényi con 100 vértices, grado promedio 50 y 10 ítems. Tanto la convicción como el factor exógeno son bajos ( $\lambda = 0.01$  y  $\alpha = 0.001$ ).

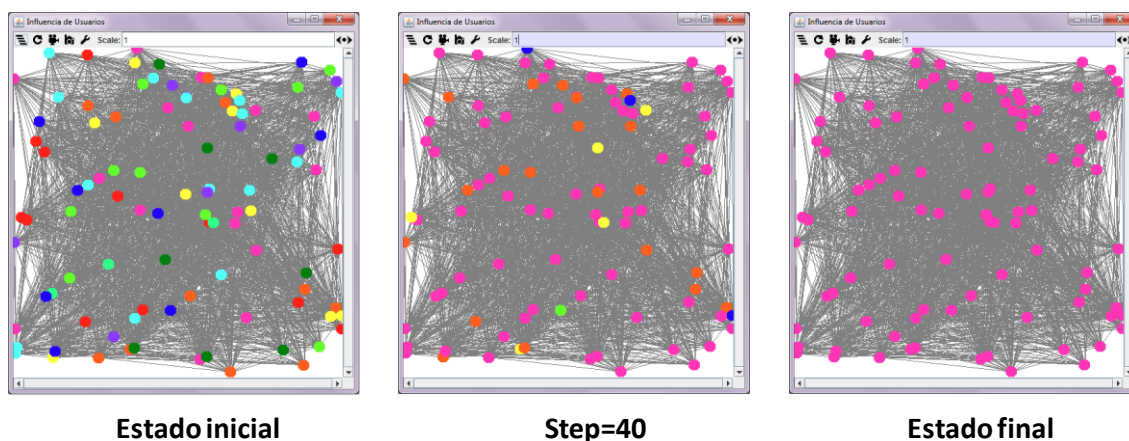


Figura 23 - Evolución preferencia por un ítem

En el estado inicial se observa la preferencia distinta de cada uno de los individuos. Esta preferencia se ve modificada en un corto espacio temporal, predominando la elección de un ítem determinado. Esta elección se extiende a todos los individuos del grafo.

Si analizamos el número de usuarios que prefieren cada ítem en cada uno de los pasos observamos cómo evoluciona en un corto espacio temporal, alcanzando el 100% de los

individuos y manteniéndose en este estado a lo largo del resto de ejecuciones, como se puede comprobar en la Figura 24.



**Figura 24 - Preferencia de ítem por usuario (caso convergencia 1 ítem)**

Este escenario suele observarse en casos en los que los individuos reciben muy poca, o casi ninguna, influencia externa o trabajamos con personas que tienen poca convicción en su opinión y, por lo tanto, son altamente influenciables por las preferencias del resto.

Cuando aproximadamente el 100% de los individuos finalicen prefiriendo el mismo ítem diremos que el sistema converge a una misma opinión.

### **Preferencia por dos ítems**

En este caso, no existe una convergencia a un ítem como en el caso anterior, sino que existe un cierto número (generalmente reducido) de ítems preferidos por los usuarios, variando su opinión entre ellos. Para el resto de ítems, no existen individuos que los prefieran.

En la Figura 25 se muestra la prueba realizada sobre un grafo Erdős-Rényi con 100 vértices, grado promedio 50 y 10 ítems, convicción  $\lambda = 0.5$  e influencia exógena  $\alpha = 0.008$ .

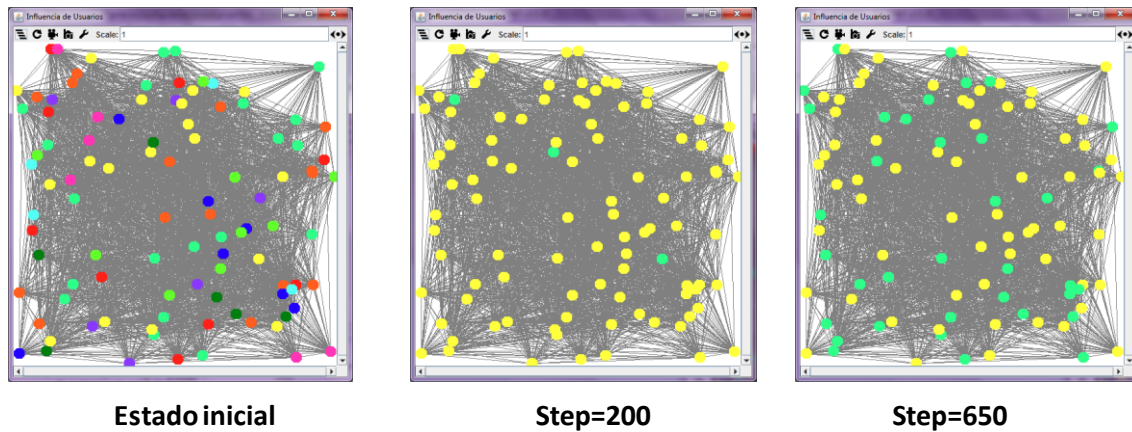


Figura 25 - Evolución preferencia 2 ítems

En la imagen se muestran tres estados obtenidos en diferentes momentos de la simulación. Al principio comienza con una distribución aleatoria de opiniones, evolucionando a un grafo con opinión casi homogénea como se observa a los 200 pasos (casi el 100% prefiere el ítem amarillo). Sin embargo vemos que, al continuar la simulación, persiste una competición entre los dos ítems (verde y amarillo).

Si observamos en la Figura 26 el número de usuarios que prefieren cada uno de los ítems en cada paso vemos cómo predomina la preferencia del ítem amarillo hasta casi el paso 500, donde el ítem verde vuelve a hacerse popular entre los usuarios.

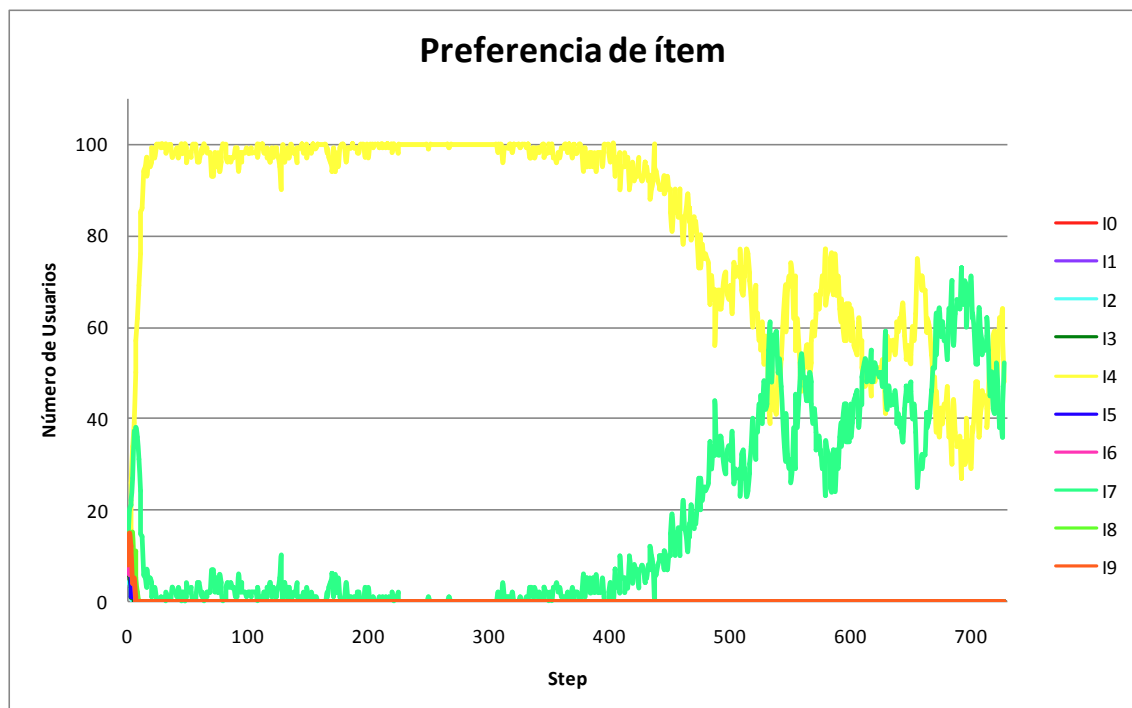


Figura 26- Preferencia de ítem por usuario (caso convergencia a 2 ítems)

A partir de ese momento, los dos ítems permanecen como los preferidos por el 100% de las personas, desapareciendo el resto.



## No convergencia

En este caso, no existe una opinión homogénea entre los usuarios del grafo. En su lugar, existen varios ítems preferidos por un porcentaje significativo de individuos. A diferencia del caso anterior, el resto de ítems no desaparecen de las preferencias de los individuos, sino que una minoría de individuos los prefiere. En la Figura 27 se muestra la prueba realizada sobre un grafo Erdős-Rényi con 100 vértices, grado promedio 50 y 10 ítems, convicción  $\lambda = 0.99$  e influencia exógena  $\alpha = 0.01$ .

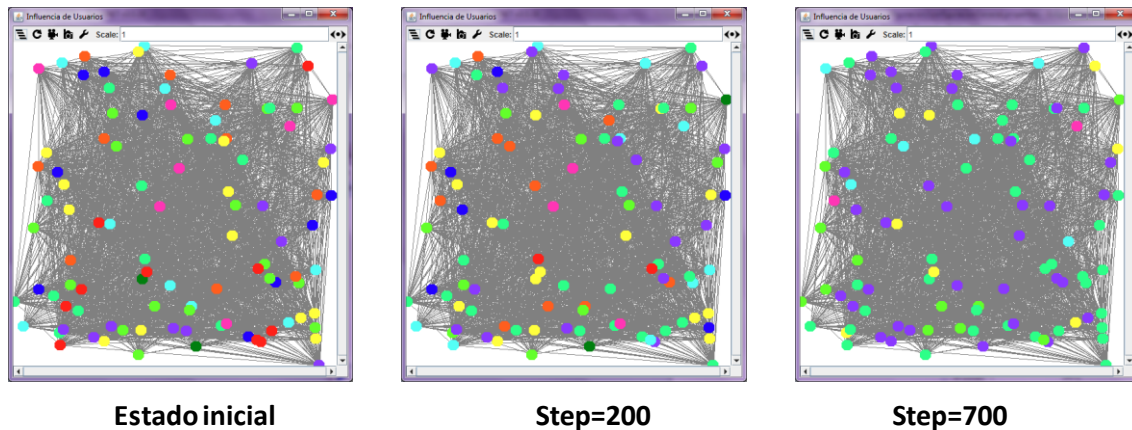


Figura 27 - Evolución no convergencia

Analizando el número de usuarios que prefieren cada ítem, no se aprecian claros ganadores. Aproximadamente el 80% de los usuarios prefieren dos ítems, sin embargo existe un 20% que se encuentran entre 4-5 ítems. Esta situación se muestra en la Figura 28.

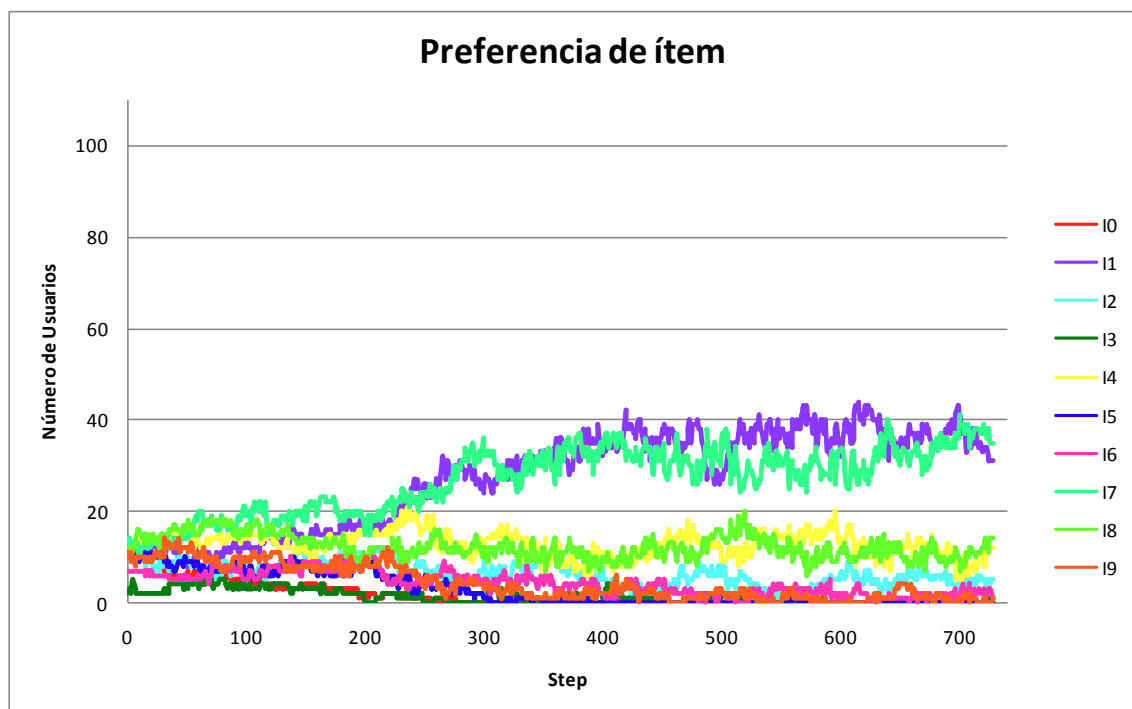


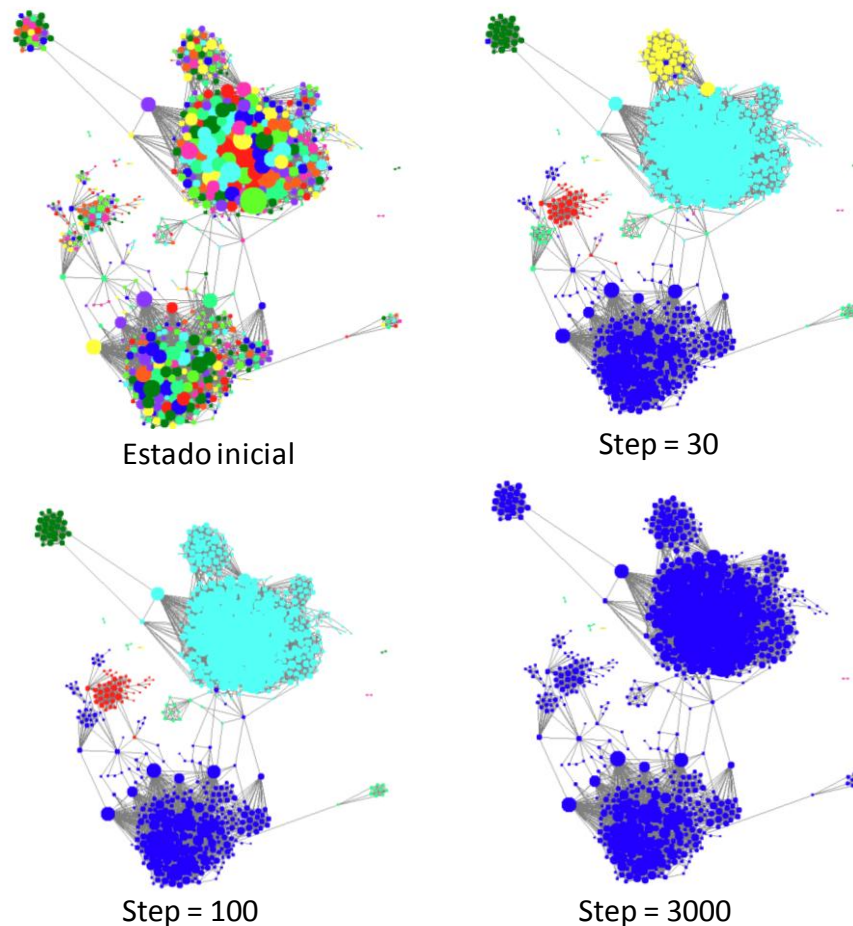
Figura 28 - Preferencia de ítem por usuario (caso no convergencia)

Para simular estos casos, hemos modificado parámetros como la convicción ( $\lambda$ ) y la cantidad de influencia exógena introducida en el sistema ( $\alpha$ ). Vamos a analizar en detalle la influencia de estos parámetros y los valores que toman.

### **Tendencias por comunidades**

El comportamiento de las redes a la propagación de influencias difiere considerablemente entre los grafos basados en modelos aleatorios y los grafos de redes sociales reales. En los grafos reales aparecen comúnmente características topológicas como la distribución de nodos en comunidades o clústers fuertemente cohesionados (y débilmente conectados con el resto del grafo), que no están presentes en los grafos Erdős-Rényi ni Barabási-Albert. En las redes altamente clusterizadas se observa cómo la evolución de la opinión tiende a ser homogénea en cada una de las comunidades.

En la Figura 29 se ejemplifica la evolución de la opinión en un subgrafo del grafo real con el que vamos a trabajar en los casos de prueba. Mostramos únicamente un subconjunto del grafo (659 nodos y 6861 aristas) para facilitar la visualización.



**Figura 29 - Evolución de la opinión en un grafo real**

En la Figura 29 podemos ver diferentes momentos de la simulación. En el estado inicial, se observa cómo los usuarios prefieren diferentes ítems. Transcurridos los 30 primeros pasos vemos cómo la opinión de cada uno de los clústers en los que se agrupan los usuarios se ha

modificado, prefiriendo todos ellos el mismo ítem. A los 100 pasos esta división se encuentra más definida, ofreciendo cada uno de los clústers resistencia a opiniones externas y que difieran de la opinión del grupo.

Finalmente, pese a las diferentes opiniones que existían, es este escenario todos los usuarios del grupo terminan prefiriendo el mismo ítem.

## 6.6 Casos de estudio

### Caso 1: Efecto de la densidad

Para ver el efecto de la densidad en la formación de opinión se han realizado pruebas sobre los dos grafos aleatorios modificando los parámetros de generación para aumentar/disminuir la densidad.

Para las pruebas se ha trabajado con valores de grado promedio de  $g(u) = \{10, 40, 80, 100, 200, 300, 400, 500, 700\}$  tanto para Erdős-Rényi como Barabási-Albert. Adicionalmente, para el modelo Erdős-Rényi se ha probado con el grafo completo  $g(u) = 1000$ .

En las pruebas se han fijado los valores intermedios tanto de influencia exógena  $\alpha = 0.005$  como de convicción  $\lambda = 0.5$ . Se ha realizado sobre 1000 nodos para poder trabajar con parámetros  $g(u)$  y  $p$  altos, evitando grafos con un gran número de aristas.

#### Erdős-Rényi

Los resultados obtenidos tras las pruebas realizadas sobre el grafo se muestran en la Figura 30.

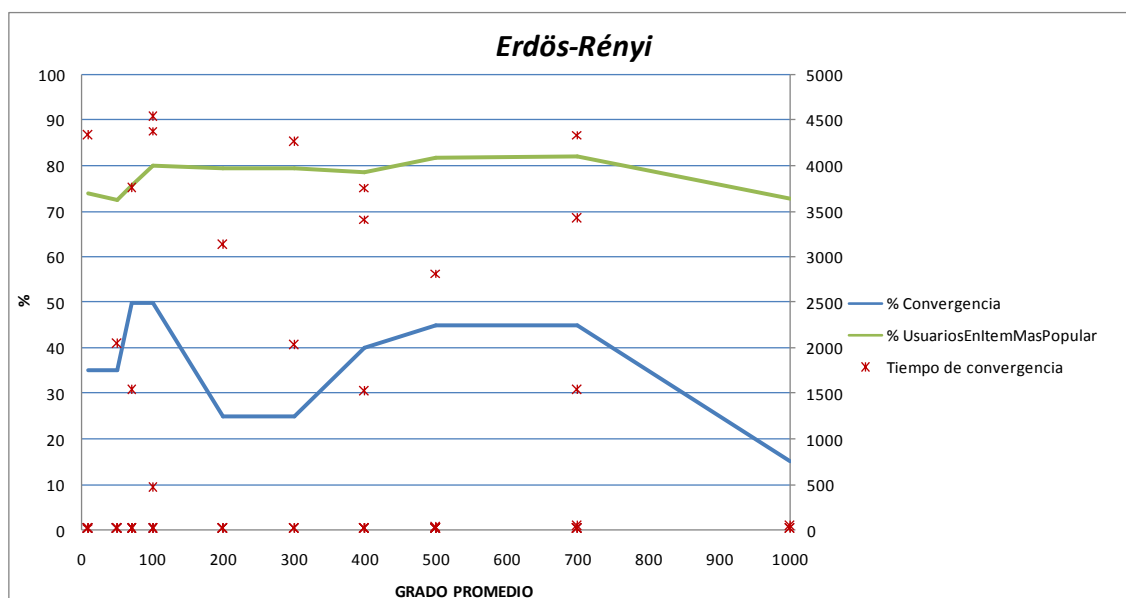
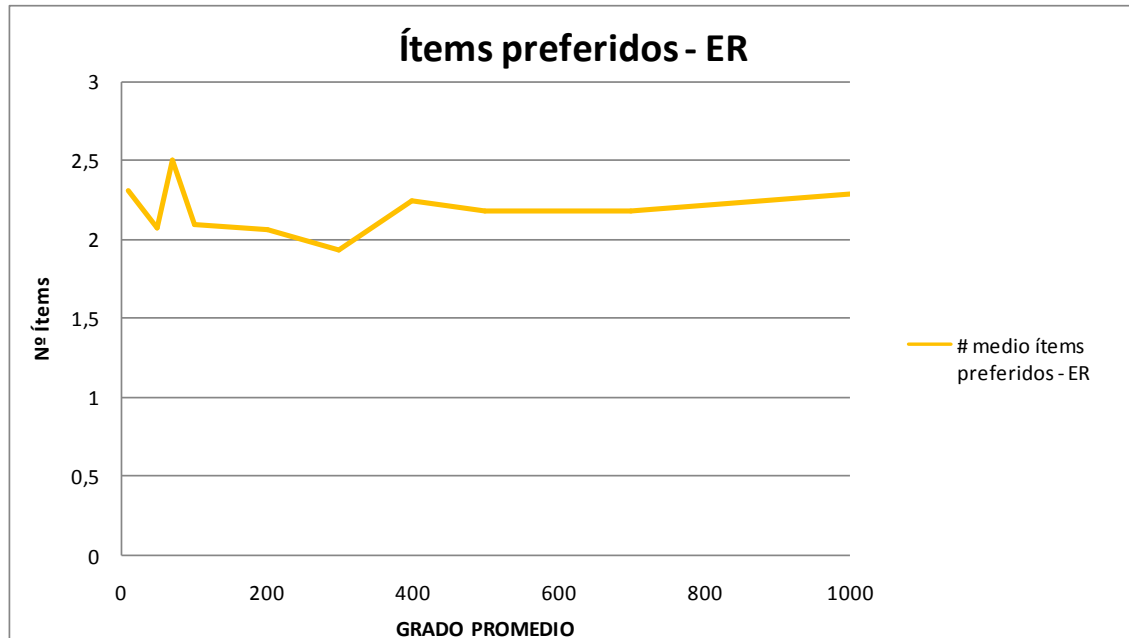


Figura 30 - Efectos observados al variar el grado promedio (Erdős-Rényi)

No parece existir relación entre la densidad del grafo y el porcentaje de convergencia, aunque sí parece disminuir el ratio de convergencia en el caso de trabajar con el grafo completo. Lo mismo ocurre con el porcentaje de usuarios que prefieren el mismo ítem en el caso de no

converger a la misma opinión; permanece alrededor del 80% aunque aumente o disminuya la densidad.

En cuanto al número de ítems preferidos en promedio, tampoco se observa una dependencia aparente como se puede ver en la Figura 31.



**Figura 31 - Número de ítems preferidos al aumentar la densidad (Erdős-Rényi)**

El número de ítems que permanecen en el grafo al finalizar la simulación permanece más o menos constante aunque se modifique la densidad.

### ***Barabási-Albert***

En este tipo de grafo ocurre lo mismo que en el anterior: aumentar la densidad del grafo no parece afectar al porcentaje de convergencia que tiene a permanecer entre un 40 y un 50% de las ejecuciones.

Lo mismo parece ocurrir al porcentaje de usuarios que prefiere en ítem más popular: en las ejecuciones que no se observa la convergencia a una misma opinión, en torno a un 80% de los usuarios prefieren el mismo ítem.

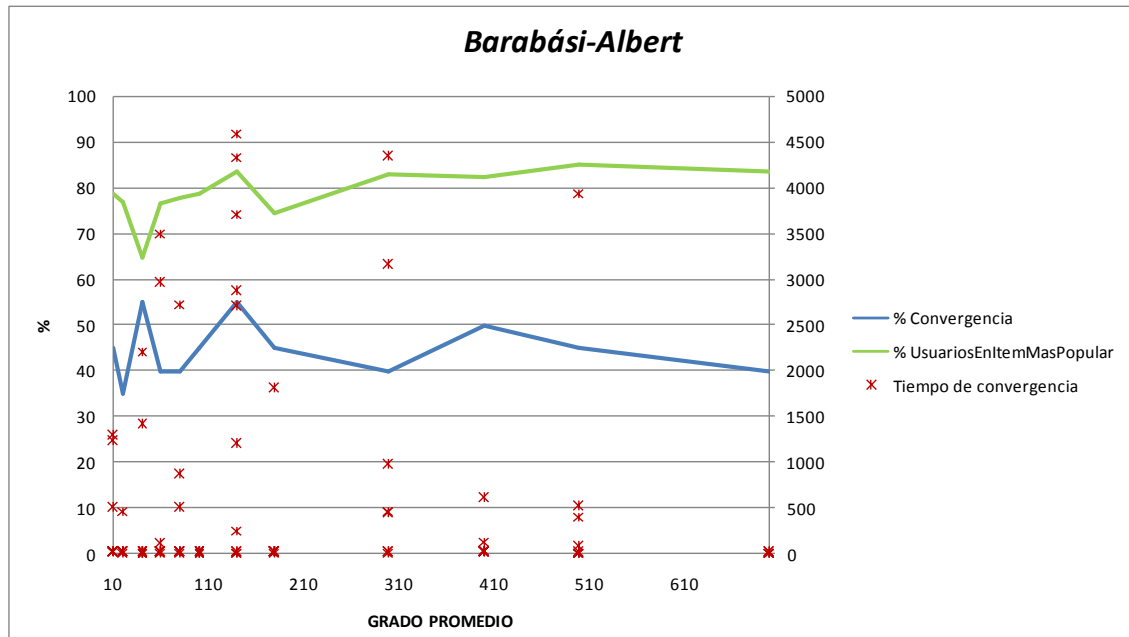


Figura 32 - Efectos observados al variar el grado (Barabási-Albert)

En cuanto al tiempo de convergencia, no parece seguir un patrón definido. Aunque para ejecuciones con grado bajo (10 -40) converge en menos de 2000 pasos y en el resto aumenta, existen excepciones como en el grado 100, con el que convergen las 9 ejecuciones en menos de 100 pasos o en el grado 400, que converge en menos de 600 pasos. Aunque este fenómeno puede haber sido casual, no se puede concluir que la densidad afecte al tiempo de convergencia.

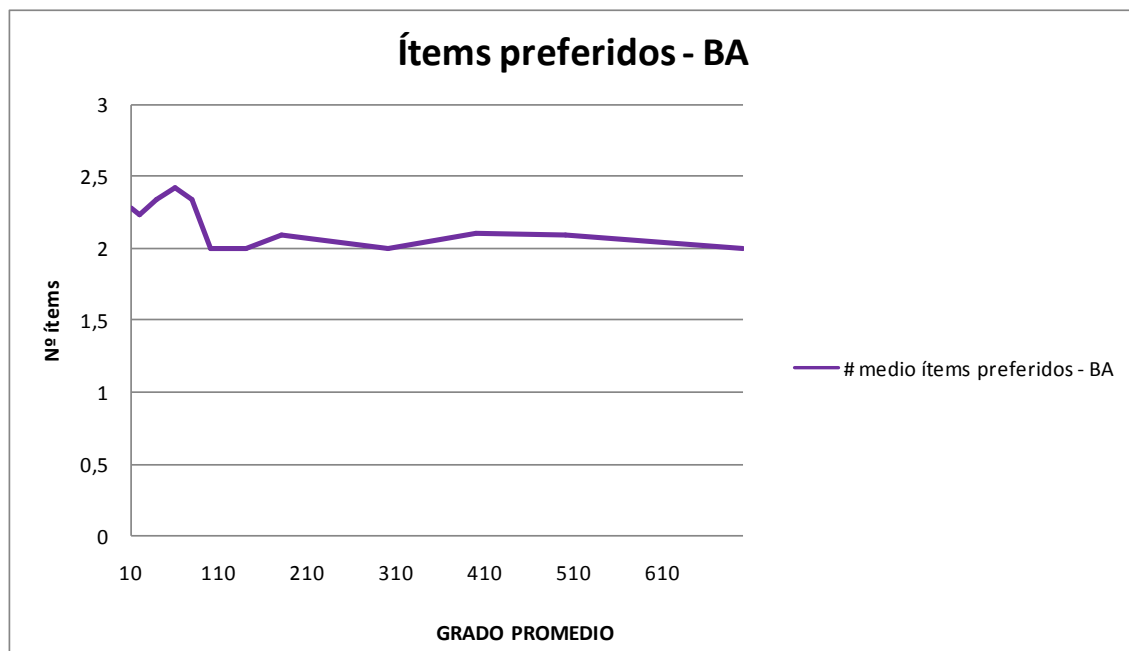


Figura 33 - Número de ítems preferidos al variar la densidad (Barabási-Albert)

En cuanto al número de ítems distintos que son elegidos por los usuarios y que se muestra en la Figura 33, parece ocurrir lo mismo que en el caso del tipo de grafo Erdős-Rényi; no se observa una relación directa con el grado promedio del grafo.

### Caso 2: Efecto de la influencia exógena

En este caso se va a comprobar el efecto del factor exógeno en el proceso de formación de una opinión. Para ello se analizan los 3 tipos de grafos: Erdős-Rényi, Barabási-Albert y el grafo real de Facebook, fijando el valor un valor de convicción intermedio ( $\lambda = 0.5$ ) y variando la influencia exógena,

$$\alpha = \{0.0, 0.001, 0.002, 0.003, 0.004, 0.005, 0.006, 0.007, 0.008, 0.009, 0.01\}.$$

Cada una de las simulaciones se ha ejecutado durante 7000 pasos.

#### Erdős-Rényi

Los resultados de las pruebas realizadas se muestran en la Figura 34.

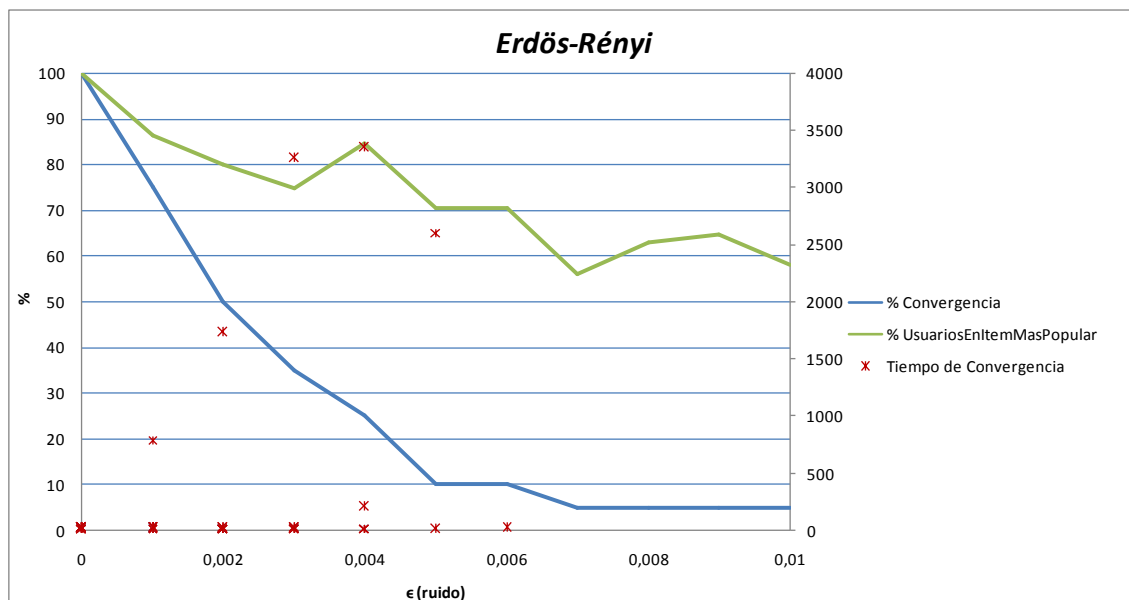


Figura 34 - Efectos observados al variar el parámetro  $\alpha$  (Erdős-Rényi)

En este tipo de grafo se observan 3 fenómenos según se introduce influencia exógena en el sistema:

- El porcentaje de ejecuciones en las que la preferencia por un ítem es predominante disminuye. Introducir influencia exógena en el sistema implica que los usuarios no acaben prefiriendo el mismo ítem.
- Lo mismo ocurre con el porcentaje de usuarios que prefieren el ítem más popular entre los usuarios del grafo: se ve cómo disminuye al aumentar el porcentaje de influencia exógena introducido. Este hecho significa que la opinión no sólo no converge a un ítem, sino que la opinión es más heterogénea.
- El paso de la ejecución en el que se produce la convergencia aumenta a partir de introducir un valor de  $\alpha$  de 0.005. Esto implica que la influencia exógena provoca que la formación de opinión se retrase.

### Barabási-Albert

En este tipo de grafo se observan resultados muy similares a los vistos en Erdős-Rényi como se puede ver en la Figura 35:

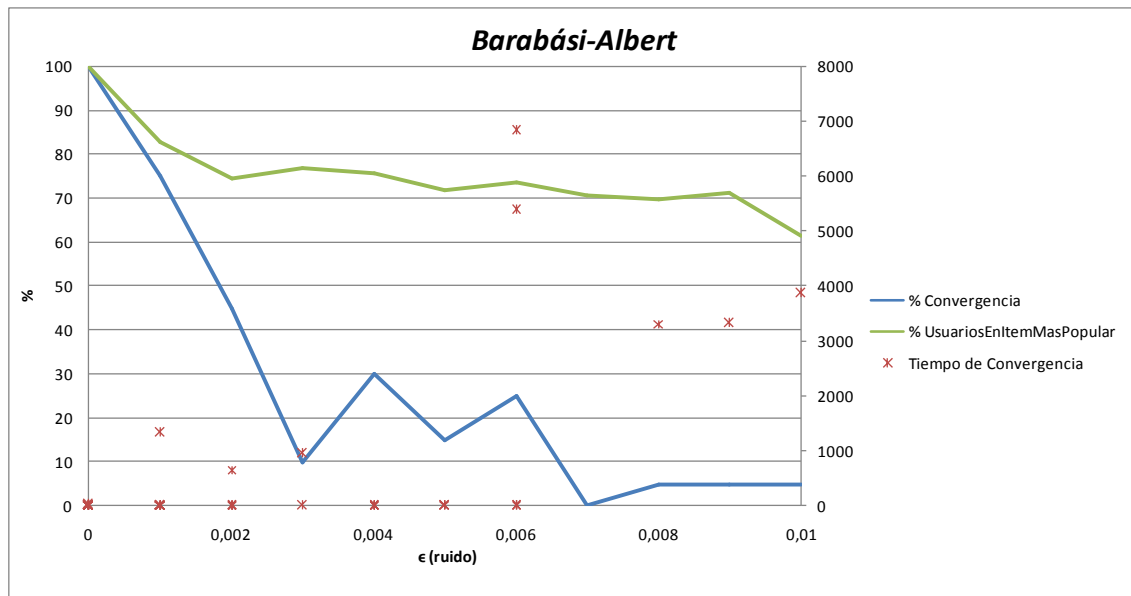


Figura 35 - Efectos observados al variar el parámetro  $\alpha$  (Barabási-Albert)

- Tanto la convergencia como el porcentaje de usuarios que prefieren el ítem mayoritario disminuyen al aumentar la introducción de influencia exógena.
- El tiempo en el cual todos los usuarios prefieren el mismo ítem aumenta cuando introducimos valores del factor exógeno más altos.

### Grafo real de Facebook

Como se puede ver en la Figura 36, en el grafo real se observan las mismas tendencias que en el resto de grafos, con algunos matices.

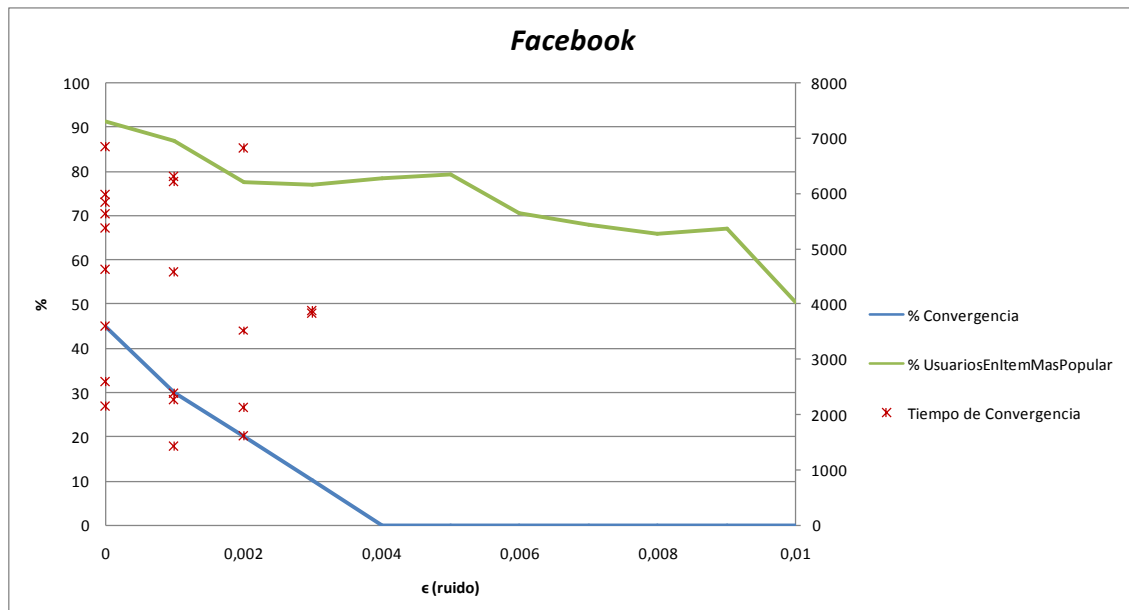


Figura 36 - Efectos observados al variar el parámetro  $\alpha$  (grafo real)

Como era de esperar, aumentar el valor del parámetro  $\alpha$  y, en consecuencia, introducir más influencia exógena disminuye el porcentaje de ejecuciones en las que todos los individuos del grafo prefieren el mismo ítem. También se aprecia la disminución del porcentaje de usuarios que prefieren el ítem mayoritario al finalizar la simulación.

Para el caso en el que no existe influencia exógena, tanto en Erdős-Rényi como en Barabási-Albert, el 100% de las ejecuciones convergían al mismo ítem. No ocurre lo mismo en el grafo real, donde sólo el 45% de las ejecuciones acaban con la misma preferencia en todos los nodos. Esta diferencia puede radicar en el número de ejecuciones realizadas puesto que si analizamos el paso en el que se llega a la preferencia uniforme aumentan de forma considerable en el grafo real.

En el caso del número de ítems que se prefieren al finalizar la simulación, en la Figura 37 se observa su crecimiento según aumenta el parámetro  $\alpha$ .



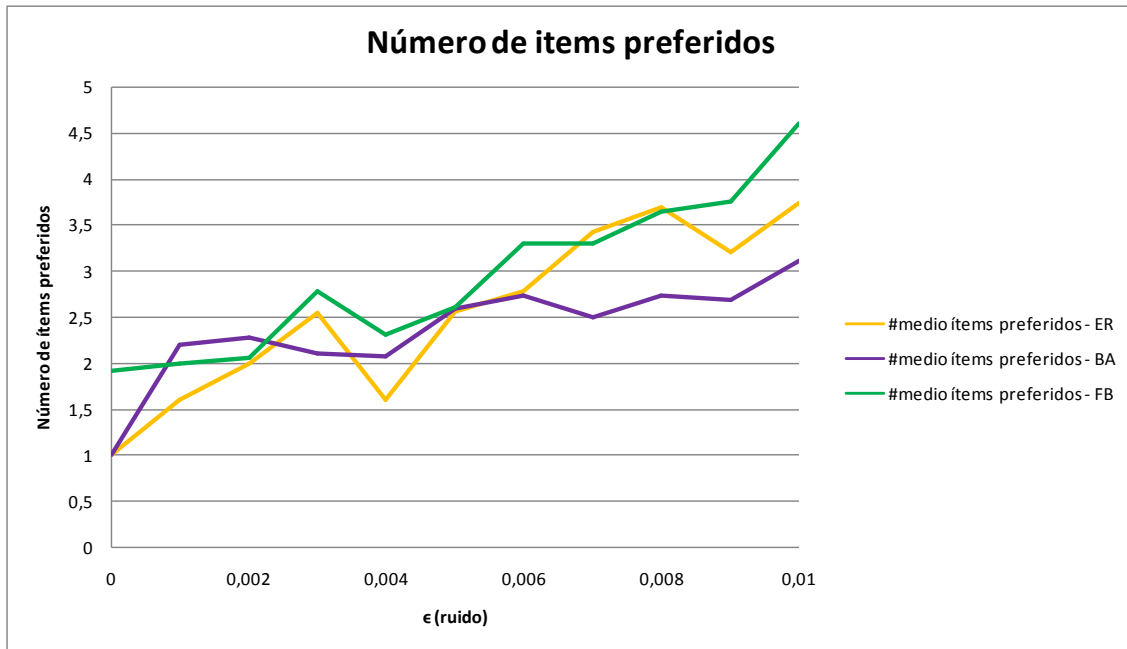


Figura 37 -Número de ítems preferidos al variar la influencia exógena

Tanto en el caso de Erdős-Rényi como Barabási-Albert, comenzaban con el 100% de ejecuciones convergiendo al mismo ítem preferido, por ese motivo en la Figura 37 en ausencia de influencia exógena ambos modelos comienzan con un único ítem. En el caso de Facebook se observa que comienza en 2 ítems cuando no existe influencia exógena puesto que, como hemos visto anteriormente, sólo el 45% de las ejecuciones convergen. En la Figura 37 se observa además cómo este número de ítems aumenta según se aumenta la influencia externa introducida.

### Caso 3: Efecto de la convicción

Al igual que en el caso anterior se analizan los 3 tipos de grafo para ver cómo influyen diferentes valores del parámetro de convicción ( $\lambda$ ) en el proceso de formación de una opinión.

Aquí fijamos el valor de influencia exógena intermedio con respecto a los valores utilizados en las pruebas del Caso 2: Efecto de la influencia exógena ( $\alpha = 0.005$ ) y variamos la convicción entre  $\lambda = 0.01$ , es decir, usuarios que no están casi nada convencidos de su opinión y se dejan influir mucho por lo que ven en el resto de sus amigos, y una convicción de  $\lambda = 0.99$ , que implicaría un caso en el que los usuarios no se dejan convencer con facilidad y tienen una convicción muy fuerte en su opinión.

#### Erdős-Rényi

La Figura 38 muestra el resultado de las pruebas ejecutadas variando los valores de convicción.

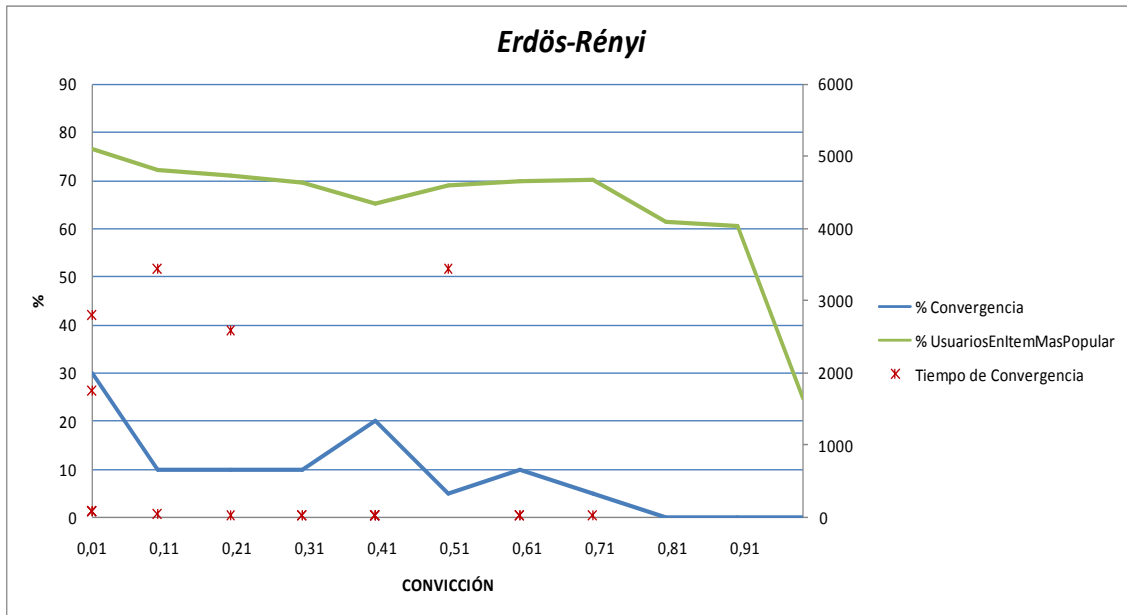


Figura 38 - Efectos observados al variar la convicción (Erdős-Rényi)

Como se observa, parece existir una disminución tanto del porcentaje de convergencia como del número de usuarios que prefieren el ítem más popular:

- Cuando la convicción es baja, los usuarios son susceptibles de ser influidos por la opinión del resto de usuarios de la red. Esta situación favorece que exista una opinión homogénea, aumentando los casos en los que todos los usuarios de la red prefieren el mismo ítem.
- Cuando aumenta la convicción, los usuarios son más resistentes a recibir influencia de su entorno y, por lo tanto, no modifican su opinión con facilidad además de hacerlo de forma muy lenta. Este fenómeno se traduce en una reducción de los casos de convergencia, hasta llegar a la situación en la que ningún caso de los 20 suceda. En la Figura 38 puede verse el aumento de ítems elegidos por los usuarios al aumentar la convicción. Con valores de  $\lambda$  cercanos a 0,99, existen más de la mitad de los ítems entre las preferencias de los usuarios. Es decir, la influencia no ha permitido modificar las opiniones para converger a las mismas opiniones.

Se observa también que el tiempo de convergencia tiende a ser bajo, quedándose generalmente en menos de 500 ejecuciones.

### Barabási-Albert

Al igual que ocurre en el ejemplo de Erdős-Rényi, aumentar la convicción en las opiniones de los usuarios provoca una disminución en el porcentaje de ejecuciones cuya opinión converge a preferir un único ítem. El mismo efecto se observa en el porcentaje de usuarios que prefieren el ítem más popular en los casos en los que no convergen a la misma opinión. Estos efectos se pueden observar en la Figura 39.

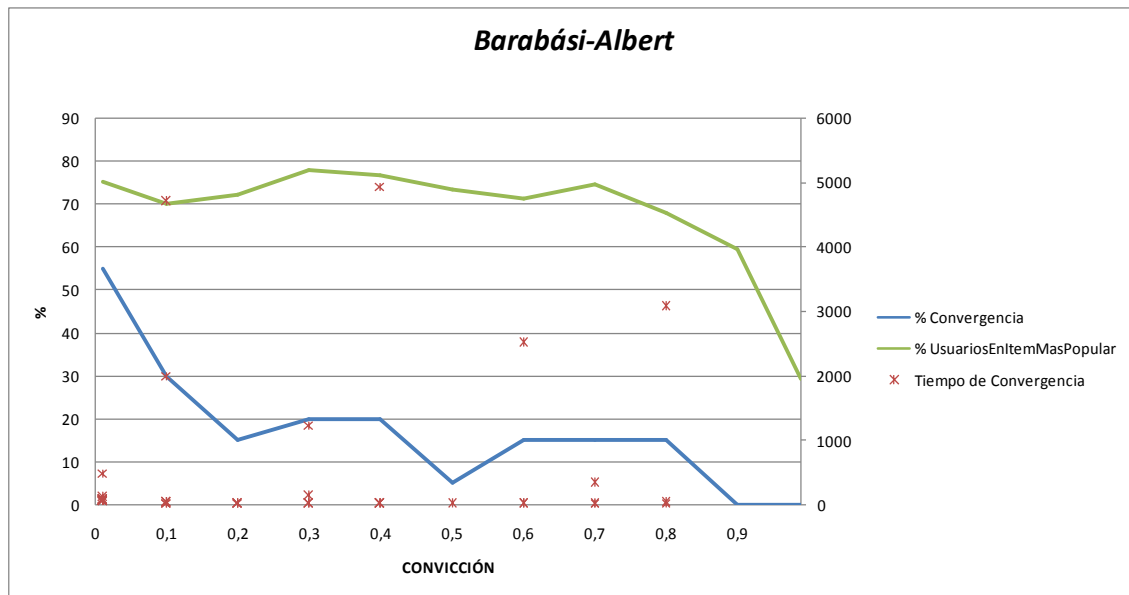


Figura 39 - Efectos observados al variar la convicción (Barabási-Albert)

En cuanto al tiempo de convergencia, parece tender a pocas ejecuciones (menos de 500) aunque existen casos en los que el tiempo de convergencia es más elevado. No parece existir relación con la convicción.

### Grafo real de Facebook

En el caso del grafo real de Facebook aunque tanto la convergencia como el porcentaje de usuarios que prefieren el ítem más popular disminuye al ir aumentando la convicción que tienen los usuarios en su opinión, el porcentaje de veces que converge a una opinión unánime es inferior al observado en los dos grafos anteriores. Incluso en las pruebas en las que la convicción es muy baja  $\lambda = 0.01$ , sólo en un 15% de las simulaciones se produce esta convergencia.

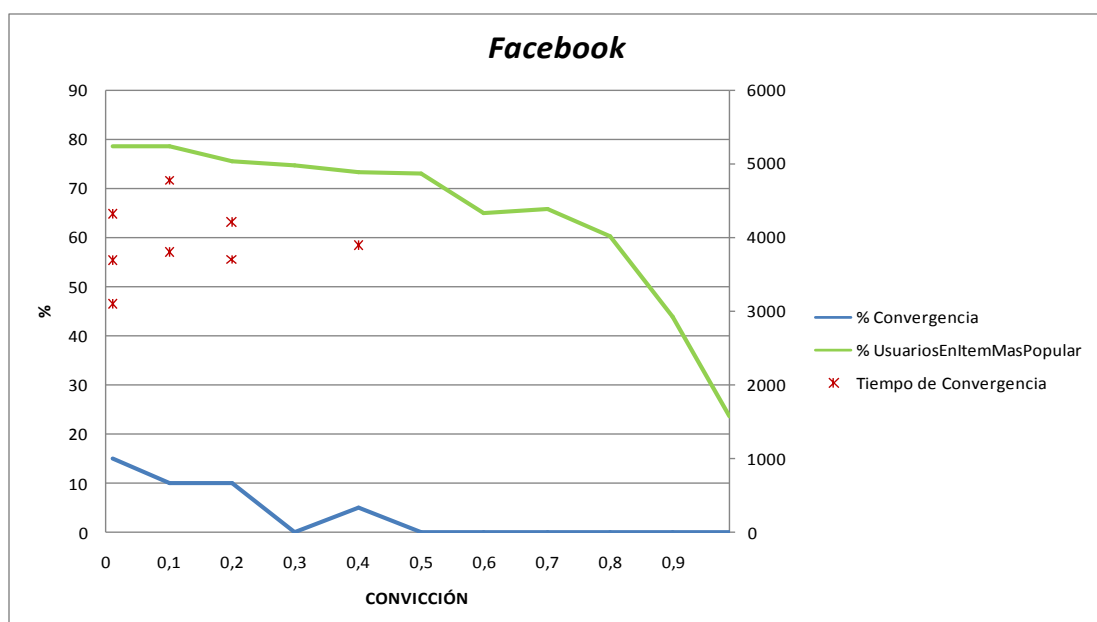


Figura 40 - Efectos observados al variar la convicción (grafo real)

También se aprecia cómo el tiempo en el que converge a una opinión homogénea ha aumentado con respecto a los dos casos anteriores.

Estas diferencias pueden ser producidas por la topología del grafo y que no es propia de los grafos Erdős-Rényi ni de Barabási-Albert, de las que ya se hablaron en la sección Tendencias por comunidades.

En cuanto al número de ítems preferidos en caso de no converger la opinión de los usuarios a preferir el mismo, se observa en la Figura 41 la relación existente con los valores de convicción  $\lambda$ : según aumenta la convicción, aumenta el número de ítems preferidos. Esto es debido a que cada usuario comienza con opiniones distribuidas de forma aleatoria y, por el alto valor de convicción, tienden a permanecer con su opinión y a ofrecer resistencia a la influencia de las opiniones del resto.

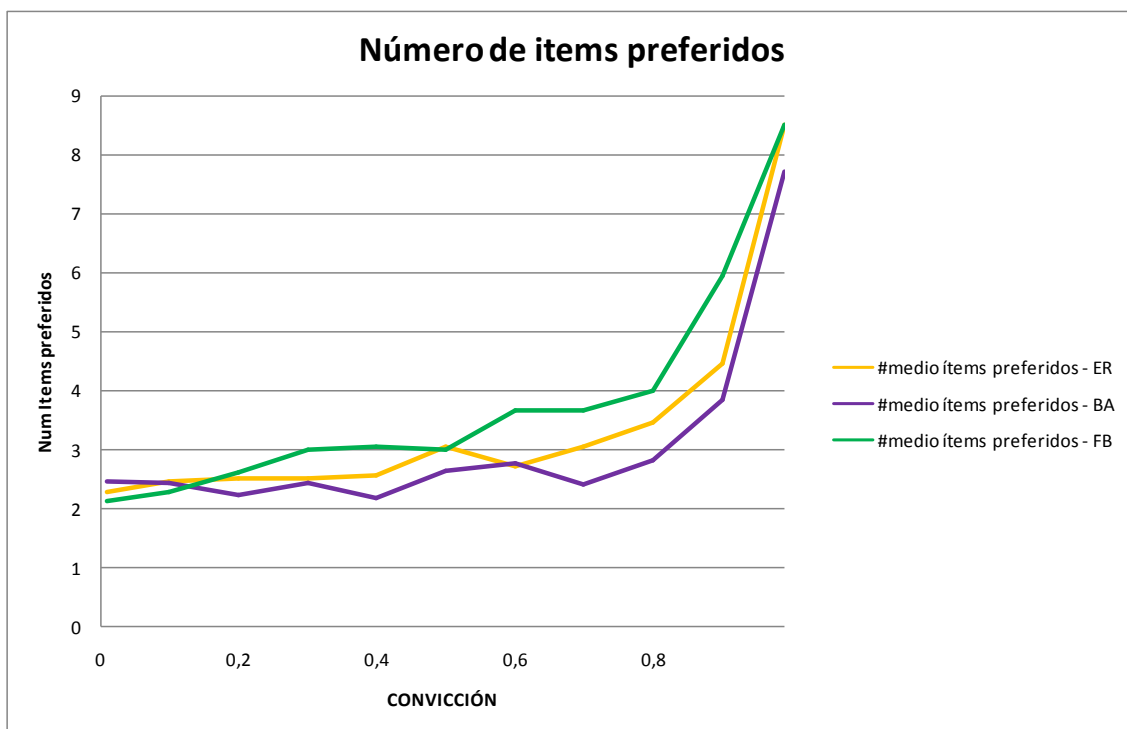


Figura 41 - Número de ítems preferidos al variar la convicción

#### Caso 4: Introducción de un agente recomendador

Para poder ver el efecto que tiene la introducción de un sistema recomendador en el modelo que presentamos, vamos a simular su introducción en el sistema. Para ello, asignamos una probabilidad del 1% de consultar la recomendación que te ofrece frente al 99% de veces restante que continuaremos consultando la opinión de nuestros vecinos.

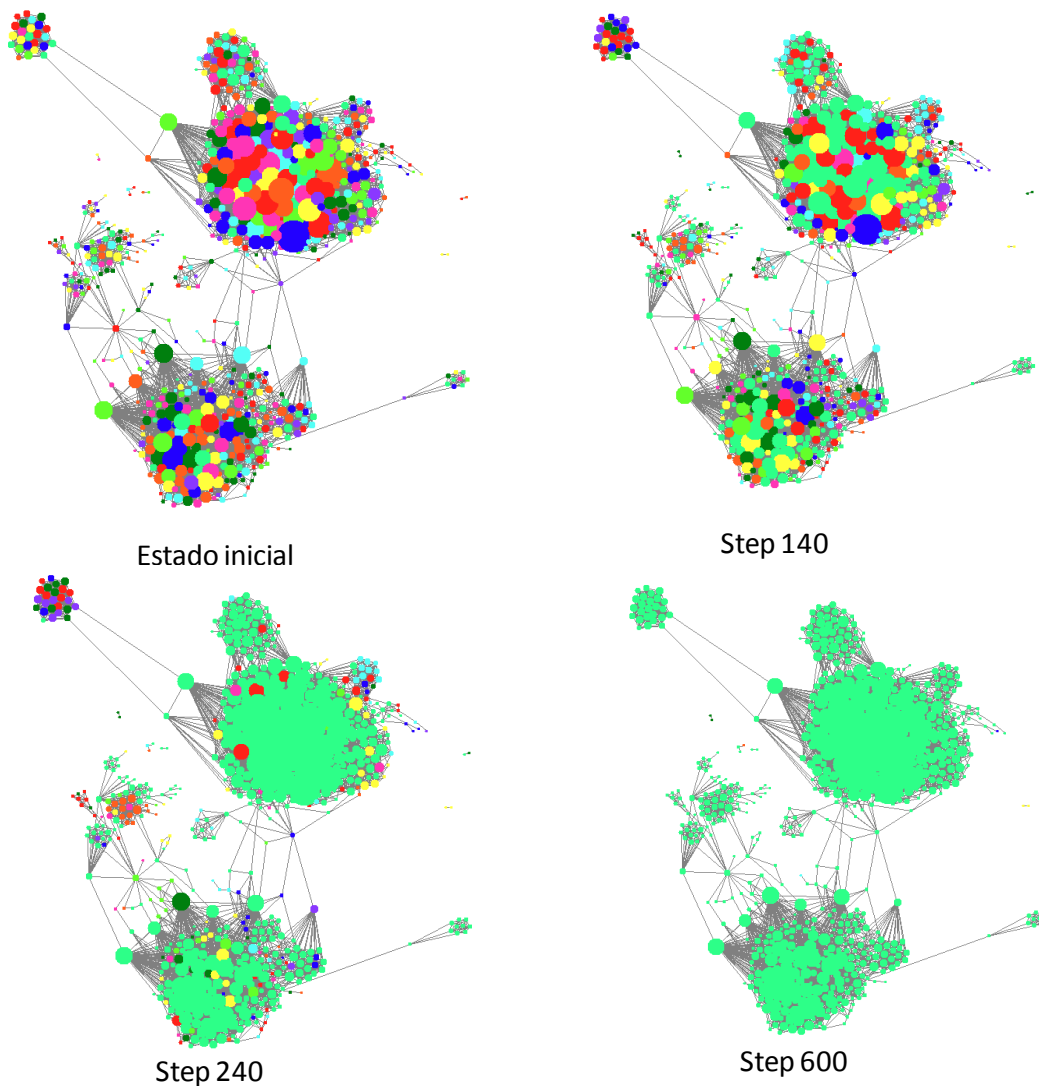
El escenario elegido es un subgrafo del grafo real de Facebook con el que hemos trabajado hasta ahora que consta de 659 nodos y 6861 aristas.

Para la elección de parámetros, hemos considerado una configuración en la que el sistema sin recomendador no converja a una opinión homogénea en la elección del ítem preferido o lo

haga en muy rara ocasión. Por este motivo, se selecciona un valor de convicción muy alto,  $\lambda = 0.99$  y vamos a variar el valor de la influencia exógena introducida en el sistema.

Aunque la interacción del recomendador sea muy baja, veremos cómo provoca un gran efecto en el sistema, produciendo una convergencia a la misma opinión en este tipo de escenarios donde no existe en condiciones normales.

En primer lugar, analizamos visualmente el efecto que produce el recomendador en la evolución de la opinión del grupo.



**Figura 42 - Evolución de la formación de opinión con recomendación**

En la Figura 42 se muestran diferentes estados de la preferencia de los individuos del grafo. En el estado inicial se observa la distribución aleatoria de opiniones. En la siguiente imagen correspondiente al paso 140, se puede ver cómo la formación de opinión no se realiza por clústers como se veía en la Figura 29, sino que se aprecia la existencia de un ítem dominante en todos ellos: el ítem con el color verde que corresponde al recomendado. Tras 240 pasos, esta opinión predomina en la mayor parte de los clústers para acabar dominando en el 100% de los individuos.

La aparición del recomendador ocasiona una ruptura del efecto observado en la formación de opinión por clústers, ya que todos tienden a la opinión del recomendador, favoreciendo la convergencia a una opinión unánime.

En la Figura 43 se muestra un escenario en el que, sin la existencia del agente recomendador, los individuos tienen preferencias distintas que no tiene por qué coincidir con las del resto de personas. En este escenario, introducimos el recomendador para ver su efecto según introducimos influencia exógena en el sistema.

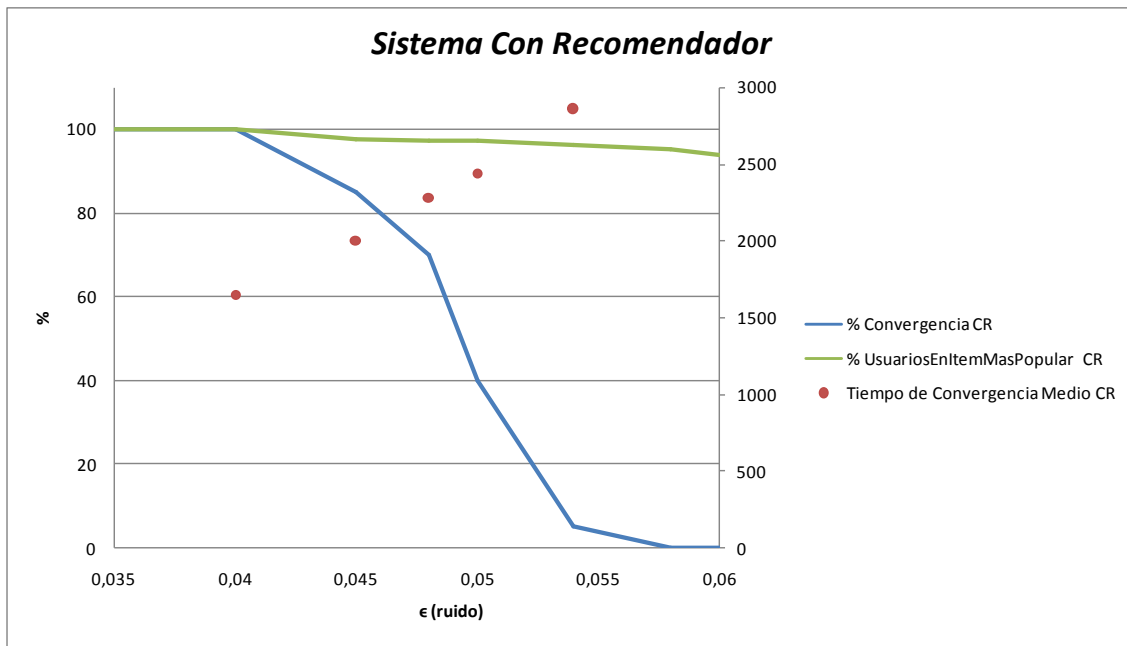
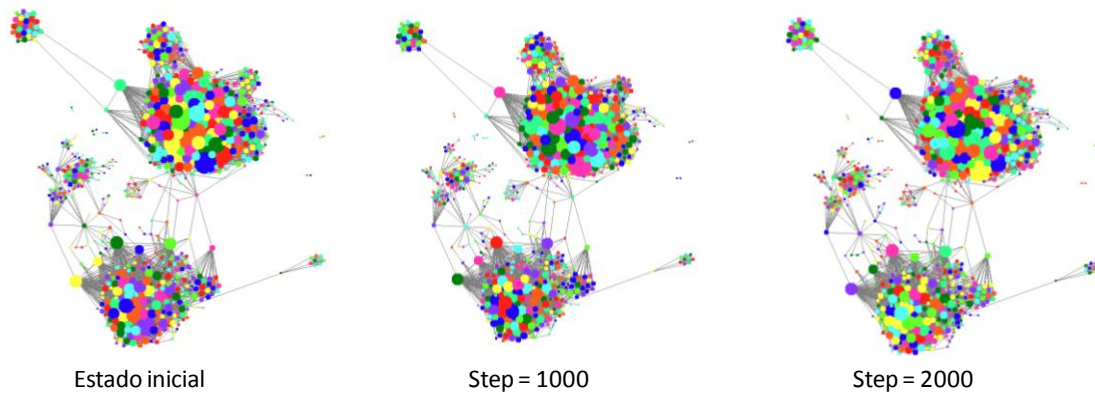


Figura 43 - Efecto del recomendador en función de la influencia exógena

En las pruebas realizadas, el efecto del recomendador provoca que todos los individuos tengan la misma opinión incluso introduciendo más influencia externa que en los escenarios anteriores.

Para valores de  $\alpha$  inferiores a 0.04 y con convicción de la opinión de cada individuo muy alta ( $\lambda = 0.99$ ), introducir el recomendador sólo en el 1% de las veces provoca que el 100% de los usuarios mantengan la misma preferencia al final de la simulación. A partir de este valor, el efecto del recomendador decae como se muestra en la gráfica.

También se observa como el tiempo en el cual todos los usuarios llegan a preferir el mismo ítem aumenta al introducir más influencia exógena en el sistema. A partir de este punto, el valor de  $\alpha$  es tan elevado que los individuos modifican su opinión constantemente. El recomendador también va modificando su recomendación, debido a que el ítem más popular fluctúa. Esta situación genera un escenario más parecido al estado inicial donde cada usuario prefiere un ítem distinto.



**Figura 44 - Efecto del recomendador con valor alto de influencia exógena**

Como se aprecia en la Figura 44, no se observan tendencias aparentes en la evolución de las preferencias de los individuos.

## **7. Conclusiones**

El presente trabajo aborda el estudio de la formación de opiniones a través de procesos de comunicación y propagación de influencia en redes sociales. Partiendo de un estudio del estado del arte al respecto, se ha desarrollado un modelo formal abstracto de propagación de influencia, que refleja escenarios reales de intercambio de opinión sobre opciones multivaluadas, con un enfoque no estudiado previamente en la literatura, hasta donde alcanza nuestra revisión del estado del arte. A partir del modelo se ha definido una dinámica de proceso que se ha implementado en un simulador y que permite observar la evolución del sistema bajo diferentes condiciones y configuraciones, tanto a nivel de tendencias cuantitativas agregadas, como de examen y visualización cualitativa. El sistema ofrece la posibilidad de modificar parámetros como la convicción o el nivel de influencia exógena, probar con diferentes tipos de grafos, tanto reales como generados por modelos aleatorios, y analizar las diferencias en los resultados.

El modelo propuesto supone un nuevo enfoque respecto a los modelos de propagación estudiados en la literatura de esta área. La introducción del concepto de multidimensionalidad en las opiniones de los ítems permite explorar un nuevo espacio en el cual se analice la preferencia de los usuarios en base a la opinión sobre diferentes objetos. Este concepto parece no haber sido abordado previamente en estudios relacionados, que o bien han considerado opiniones unidimensionales (Hammer, Moore, Finley, & Glass, 2012), o bien estados binarios (p.e. infectado, no infectado (Newman M. , 2010)), o bien consideran la multidimensionalidad en otro contexto (Salganik, Dodds, & Watts, 2006).

El presente estudio busca contribuir a determinar y comprender mejor en qué forma y medida los parámetros que caracterizan la topología de una red social, la distribución del poder de convicción de los usuarios, su receptividad, su comportamiento en la comunicación, etc., a nivel individual, influyen en las derivas globales del estado de la red, su posible convergencia hacia situaciones estacionarias características, la velocidad de cambio, el grado de pluralidad y estabilidad de la opinión general, y otros patrones característicos en la evolución de un sistema de preferencia y elección de alternativas en sociedades interconectadas. Para ello se han medido aspectos como el porcentaje de individuos que acaban eligiendo el mismo ítem al final de las ejecuciones o el tiempo en el cual se produce este fenómeno en función de diferentes parámetros como la densidad del grafo, el nivel de convicción de los usuarios en su propia opinión o la influencia exógena.

Se han identificado tres casos principales en la evolución del sistema definido:

- Aquellos en los que emerge una preferencia homogénea, es decir, todos los individuos del grafo acababan prefiriendo el mismo ítem.
- Aquellos en los que coexiste el predominio de un pequeño número de opiniones que se estabilizan destacando del total inicial (10 ítems en total en nuestras pruebas), en tanto que el resto de ítems se extinguen de entre los preferidos por los usuarios.
- Aquellos en los que no emergen consensos y los individuos prefieren ítems distintos sin estabilizarse ninguna tendencia marcada.



Además, en el caso de aplicar el modelo sobre grafos reales donde encontramos la presencia de clústers, observamos cómo se produce la formación de opinión de forma aislada por comunidades.

A nivel más estadístico, hemos estudiado dependencias entre condiciones de configuración del sistema (densidad de la red, convicción de los usuarios y nivel de influencia exógena) y características de la evolución del sistema. Las observaciones obtenidas son las siguientes:

- Según aumenta el nivel de convicción de los usuarios, más dificultad encuentra el sistema en llegar a un consenso en las preferencias de los individuos. Valores bajos de convicción implican que los individuos son altamente influenciables por lo que, aunque se introduzca algo de influencia exógena en el sistema, acaban prefiriendo el mismo ítem en aproximadamente la mitad de las simulaciones realizadas en nuestro estudio. A medida que aumenta la convicción, disminuye esta influencia y, por lo tanto, tiene más peso la opinión individual. Este hecho ocasiona un descenso en el número de veces que los usuarios convergen a una preferencia unánime.
- Al igual que ocurre con la convicción, al aumentar el valor del parámetro  $\alpha$  y, en consecuencia, introducir más influencia exógena en el sistema, el número de ejecuciones en las que el sistema converge disminuye. En las pruebas realizadas con parámetro de convicción media ( $\lambda = 0.5$ ), en ausencia de influencia exógena el sistema converge a una preferencia por el mismo ítem común en el 100% de las simulaciones.
- Por último, al aumentar el grado promedio en los grafos tanto de Erdős-Rényi como de Barabási-Albert no parece observarse un efecto en el porcentaje de veces que el sistema converge a un estado con preferencia común.

Junto con estos experimentos, se ha realizado una prueba para analizar el efecto de la introducción de un agente recomendador en el sistema, concretamente un método basado en popularidad, que recomienda a los usuarios el ítem preferido por el mayor número de individuos en cada momento. Se observa un drástico incremento en el número de veces que la población converge a una misma preferencia, además de la ruptura de formación de opiniones por clúster en el grafo real.

## **8. Trabajo futuro**

El trabajo realizado en este TFM abre numerosas vías de continuación, tanto a nivel del modelo y sus múltiples posibilidades de extensión, como del programa de simulación que se ha implementado. Se apuntan a continuación algunas de ellas.

En la observación cualitativa del proceso de formación de opinión en grafos reales, hemos observado comunidades cuya resistencia a preferir el ítem de la mayoría es mayor que el resto. Queda pendiente medir las características estructurales de las comunidades tales como el coeficiente de clustering, el grado promedio o la densidad con el fin de ver si existe correlación, así como el posible papel de individuos y conexiones especiales en la red, tales como hubs (usuarios altamente conectados), brokers (puntos de paso), y los denominados enlaces débiles entre comunidades.

Tiene interés asimismo ampliar el rango de grafos con los que experimentar, en particular con datos reales de colecciones públicas, redes sociales de diferente naturaleza (p.e. red de interacción en Twitter). En esta línea, se podrían considerar aspectos más realistas en el modelo, como el hecho de que la red social no es estática y que su formación y origen están relacionados con la propia interacción de la que después van a ser vehículo (dicho de otro modo los contactos se forman a través de la interacción). Por otro lado no interactuamos con la misma frecuencia con todos nuestros conocidos, y esto se podría reflejar ponderando los arcos de forma que la selección aleatoria de vecinos sea con arreglo a estos pesos.

En esta misma línea, sería natural combinar nuestro modelo con el concepto de tolerancia introducido por Hammer *et al.* (2012), limitando la interacción de un individuo con los más similares. Además, introducir el concepto del líder como aquel usuario cuya opinión es fuerte y no recibe influencia de la opinión del resto de individuos. Sería equivalente al concepto de inmunidad en modelos de epidemias, introduciendo usuarios ‘blindados’ ante la influencia del resto. El concepto se puede generalizar a una distribución no uniforme de la convicción y la persuasividad.

Igualmente en el campo de los grafos aleatorios, se puede experimentar con distribuciones del grado con diferente grado de sesgo, desde una distribución uniforme hasta grafos más abruptos incluso que el modelo de Barabási-Albert. Estas distribuciones arbitrarias se pueden formar con los llamados modelos de configuración, indicando la secuencia deseada de grados, a fin de estudiar el efecto del sesgo en el comportamiento del sistema.

Cabe asimismo explorar múltiples combinaciones de aproximaciones alternativas a las del modelo propuesto, en el orden y secuencia de interacciones entre usuarios y las dimensiones de opinión. Se podría por ejemplo modificar en cada instante de tiempo  $t$  una única opinión de un ítem escogido al azar, o interactuar con una persona única para cada ítem (en lugar de un vecino diferente para cada ítem). Cada usuario podría comunicarse en su turno de interacción con todos sus vecinos para todos los ítems. Los turnos podrían ser aleatorios en lugar de exhaustivos (una iteración completa sobre todos los usuarios en el bucle externo de la simulación). La conversación entre dos usuarios se podría limitar al ítem preferido por cada uno, en lugar de un ítem aleatorio o todos los ítems. Y así sucesivamente. Se pueden contemplar asimismo diversas variantes en la definición de las condiciones iniciales. Por ejemplo, las opiniones iniciales podrían distribuirse de forma no aleatoria, basada en preferencias por un ítem concreto por un conjunto de individuos de la red o por clúster, o cada usuario podría partir de un único ítem con preferencia positiva, etc.

Puede igualmente generalizarse la influencia exógena de forma que, por ejemplo, no sea necesariamente uniforme respecto a los ítems como se ha configurado aquí. Una distribución sesgada sobre los ítems representaría situaciones como las campañas de marketing sobre los ítems. Bajo este mismo concepto pueden verse asimismo los sistemas de recomendación como un caso particular de factor exógeno a la red, de particular interés, ya que los recomendadores no sólo influyen en las opciones y preferencias que se forman los usuarios, sino que beben de éstas, creando así un ciclo complejo de retroalimentación interesante de investigar. Se abre en esa dirección un espacio amplio a explorar, donde múltiples estrategias de recomendación pueden tener efectos diversos.

## **9. Bibliografía**

- Adomavicius, G., & Tuzhilin, A. (2005). Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. *IEEE Trans.on Knowledge and Data Engineering* 17 (6), (págs. 734-749).
- Bakshy, E., Rosenn, I., Marlow, C., & Adamic, L. (2012). The Role of Social Networks in Information Diffusion. *Proceedings of the 21st international conference on World Wide Web*. ACM, (págs. 519-528). Lyon, France.
- Barabási, A. -L., & Albert, R. (1999). Emergence of scaling in random networks. *Science* , 286 (5439), 509-512.
- Bearden, J., Atwood, W., Freitag, P., Hendricks, C., Mintz, B., & Schwartz, M. (1975). The nature and extent of bank centrality in corporate networks. *Annual Meetings of the American Sociological Association*.
- Bonchi, F. (2011). Influence Propagation in Social Networks: A Data Mining Perspective. *IEEE Intelligent Informatics Bulletin* , 12 (1), 8-16.
- Bott, E. (1955). Urban families: Conjugal roles and social networks. *Human Relations* , 8, 345-84.
- Boyd, D. M., & Ellison, N. B. (2008). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication* , 210-230.
- Burt, R. S. (1982). *Towards a Structural Theory of Action*. Academic Press .
- Chandra, A. K. (2012). Percolation in a kinetic opinion exchange model. *Physical Review E* , 85 (2), 021149.
- Doerr, B., Fouz, M., & Friedrich, T. (2012). Why Rumors Spread Fast in Social Networks. *Communications of the ACM* .
- Domingos, P., & Richardson, M. (2001). Mining the network value of customers. *Proc. of the Seventh ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD'01)*, (págs. 57-66). San Francisco, CA, USA.
- Erdős, P., & Rényi, A. (1959). On random graphs. *Publicationes Mathematicae Debrecen* , 6, 290-297.
- Goyal, A., Bonchi, F., & Lakshmanan, L. V. (2010). Learning Influence Probabilities In Social Networks. *Third ACM international conference on Web search and data mining (WSDM 2010)*, (págs. 241-250). New York City, NY, USA .
- Hammer, R. J., Moore, T. W., Finley, P. D., & Glass, R. J. (2012). The Role of Community Structure in Opinion Cluster Formation. *2nd International Conference on Complex Sciences: Theory and Applications*. Santa Fe, New Mexico.
- Kempe, D., Kleinberg, J., & Tardos, É. (2003). Maximizing the Spread of Influence through a Social Network. *9th ACM international conference on Knowledge discovery and data mining (SIGKDD 2003)*, (págs. 137-146). New York, NY, USA.
- Kleinberg, J., & Easley, D. (2010). *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press.

- Lallouache, M., Chakrabarti, A. S., Chakraborti, A., & Chakrabarti, B. K. (2010). Opinion formation in the kinetic exchange models: Spontaneous symmetry breaking. (A. P. Society, Ed.) Phys. Rev. E , 82, 056112.
- Liu, F., & Lee, H. J. (2010). Use of social network information to enhance collaborative filtering performance. Expert Systems with Applications: An International Journal , 37 (7), 4772–4778.
- McAuley, J., & Leskovec, J. (2012). Learning to discover social circles in ego networks. Advances in Neural Information Processing Systems 25 , 548-556.
- Milgram, S. (1967). The small-world problem. Psychology Today , 2 (1), 60-67.
- Newman, M. E. (2005). Power laws, Pareto distributions and Zipf's law. Contemporary physics , 46 (5), 323-351.
- Newman, M. (2010). Networks An Introduction. Oxford University Press.
- Salganik, M. J., Dodds, P. S., & Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. science , 311 (5762), 854-856.
- Scott, J., & Carrington, P. J. (Edits.). (2011). The SAGE Handbook of Social Network Analysis.
- Shakarian, P., & Paulo, D. (2012). Large Social Networks can be Targeted for Viral Marketing with small seed sets. Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012), (págs. 1-8). Kadir Has University, Istanbul, Turkey.
- Sinha, R., & Swearingen, K. (2001). Comparing recommendations made by online Systems and Friends. DELOS workshop: personalisation and recommender systems in digital libraries, 106.
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of "small-world" networks. Nature , 393, 440-42.
- Wu, F., & Huberman, B. A. (2004). Social structure and opinion formation. arXiv preprint cond-mat/0407252 .